Benzoic acid metabolism in *Sorbus aucuparia* cell suspension cultures

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina

zu Braunschweig

zur Erlangung des Grades einer

Doktorin der Naturwissenschaften

(Dr.rer.nat.)

genehmigte

D i s s e r t a t i o n

von Mariam Mamdoh Magdy Nashed Gaid
aus Assiut / Ägypten

1. Referent: Professor Dr. Ludger Beerhues
2. Referent: apl. Professor Dr. Dirk Selmar
eingereicht am: 16.06.2010
mündliche Prüfung (Disputation) am: 17.09.2010

Druckjar 2010

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakulät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Gaid MM, Sircar D, Beuerle T, Mitra A, Beerhues L. Benzaldehyde dehydrogenase from chitosan-treated *Sorbus aucuparia* cell cultures. *J Plant Physiol* 166: 1343-1349 (2009)

Gaid MM, Scharnhop H, Ramadan H, Beuerle T, Beerhues L. 4-Coumarate:CoA ligase family members from elicitor-treated *Sorbus aucuparia* cell cultures producing benzoate-derived phytoalexins. Submitted **(2010)**

Tagungsbeiträge

Mariam M Gaid, Debabrata Sircar, Till Beuerle, Adinpunya Mitra, Ludger Beerhues; Benzoic acid metabolism in *Sorbus aucuparia* cell cultures (Poster). Deutsche Botanische Gesellschaft, Botanikertagung, University of Leipzig, 6-10th September, 2009

Ines Belhadj, **Mariam M Gaid**, Ludger Beerhues; Hyperforin biosynthesis: cDNA cloning of isobutyrophenone synthase (Poster), 58th International Congress and Annual Meeting of the Society for Medicinal Plant and Natural Product research, Berlin, 29th August- 2nd September 2010

ACKNOWLEDGMENT

I would like to express my special gratitude to my supervisor Professor Dr. *Ludger Beerhues* for providing the interesting research point and facilities to pursue my PhD study at the Institut für Pharmazeutische Biologie. His valuable supervision, guidance, enthusiasm and especially great patience have facilitated the completion of this dissertation. It is important to mention that four years ago I was feeling my way through my project. There were days when I was trying to figure out the problems, which can be quite frustrating for someone new. But you've helped me through that. Now, I look back on that time with a bit of amusement. The things that seemed so difficult then are a breeze now.

I'm also thankful to Professor Dr. *Benye Liu* whose patience, guidance, constructive discussion, valuable ideas and support enabled me to finish this work.

My great appreciation is given to Dr. *Till Beuerle* for his GC-MS and LC-MS analyses, analytical assistance and useful discussions.

I would like to express my appreciation to Dr. *Rainer Lindigkeit* for much helpful advice with software problems.

My warmest thanks to my colleagues of the Institute of Pharmaceutical Biology at the TU-Braunschweig, present or past, who made this study a successful experience. I especially thank *Ines Rahaus* for sequencing my DNA samples and her timely help. Dr. *Asya Swiddan, Rawad Zodi, Andreas Müller, Iman Abd-El Rahman, Ines Belhadj, Maren Lütge* and *Doris Glindemann* for the pleasant working environment and companionship during laboratory works.

I'm so grateful to *Cornelia Hüttner* for her timeless help during my stay in Germany, encouragement and her presence when needed. Honestly, I shall never find words to thank her.

I'm very lucky to have a senior colleague like Dr. *Helge Scharnhop* in our research group. It is my privilege to sincerely thank him for his continuous help in all aspects.

My special thanks go to my colleague Dr. *Debabrata Sircar*. I really appreciate his encouragement and support. If I can ever return the favor, please let me know. I would gladly be of help.

I'm grateful to my colleague *Inis Winde* for her help in radioisotopic assays.

I acknowledge the *Egyptian Government* for granting me a PhD-scholarship.

My deep thanks to the staff members of my department in Egypt for offering me such an opportunity to study in Germany and for their continuous encouragement.

I would like to express my special thanks to my husband *Hany* for his support and love and to my son *Kerolos* for his lovely smile that can relieve any kind of tiredness.

Finally, my sincere thanks are given to my family in Egypt, especially my mother Professor Dr. *Sozan Tous* (Department of Pharmaceutics, Assiut University, Egypt) who provided me with continuous moral support during my stay in Germany and my whole life.

List of abbreviations

2,4-D	2,4-Dichlorophenoxyacetic acid
4CL	4-Coumarate-coenzyme A ligase
6x His	Hexa histidine tag
A	Ampere
ALDH	Aldehyde dehydrogenase
AMP	Adenosine-5`-monophosphate
Amu	Atomic mass unit
APS	Ammonium persulphate
ATP	Adenosine triphosphate
Au	Absorption Unit
BD	Benzaldehyde dehydrogenase
BIS	Biphenyl synthase
bp	Base pair
BPS	Benzophenone synthase
BSA	Bovine serum albumin
BZL	Benzoate-CoA ligase
C4H	Cinnamate 4-hydroxylase
cDNA	Complementary deoxyribonucleic acid
CHS	Chalcone synthase
CNL	Cinnamate-CoA ligase
CoA	Coenzyme A
cpm	Count per minute
ddNTP	2´,3´-Dideoxynucleoside triphosphate (terminator nucleotide)
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
dsDNA	Double strand DNA
DTT	1,4-Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylendiamine-tetraacetic acid
EPI	Enhanced product ion scan
ESI-MS	Electrospray ionization- mass spectrometry
EtOH	Ethanol
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
GC-MS	Gas chromatography-mass spectrometry
GSP	Gene specific primer
GST	Glutathione-S-transferase
h	hour
HEPES	Hydroxyethylpiperazin ethan sulphonic acid
HPLC	High performance liqiud chromatography
HPSF	High Purity Salt Free
IPTG	Isopropyl-β-D-thiogalactopyranoside
KDa	Kilo Dalton
K_m	Michaelis-Menten constant
kp	Kilo base pair
L	Liter
LB	Luria Bertani
L-Phe	L-Phenylalanine
LS	Linsmaier and Skoog
MCS	Multiple cloning site
MeOH	Methanol
Min	Minute
M-MuLV	Moloney Murine Leukemia Virus
MWCO	Molecular weight cutoff
NAA	1-Naphthaleneacetic acid

NAD^+	Nicotinamide adenine dinucleotide
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
Ni-NTA	Nickel-nitrilotriacetic acid
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAL	Phenylalanine ammonia-lyase
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Pfu	*Pyrococcus furiosus*
PKSs	Polyketide synthases
RACE	Rapid Amplification of cDNA Ends
rpm	Revolution per minute
RT	Reverse transcriptase
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulphate
Sec.	Second
SMART	Switching Mechanism At 5` end of RNA Template
Taq	*Thermus aquaticus*
TD primer	TATA-box degenerate primer
TEMED	*N,N,N',N'*- tetramethylethylene diamine
Tris	Tris-(hydroxymethyl)aminomethan
U	Unit
X-Gal	5-Bromo-4-chloro-3-indolyl-*β*-D-galactopyranoside
ε	Extinction coefficient

Amino acids	
A=Ala	Alanine
C=Cys	Cysteine
D=Asp	Aspartic acid
E=Glu	Glutamic acid
F=Phe	Phenylalanine
G=Gly	Glycine
H=His	Histidine
I=Ile	Isoleucine
K=Lys	Lysine
L=Leu	Leucine
M=Met	Methionine
N=Asn	Asparagine
P=Pro	Proline
Q=Gln	Glutamine
R=Arg	Arginine
S=Ser	Serine
T=Thr	Threonine
V=Val	Valine
W=Trp	Tryptophan
Y=Tyr	Tyrosine

Nucleotides	
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
I	Inosine

I. Introduction

1 *Sorbus aucuparia* L.

The plant genus *Sorbus* comprises about 100–200 species of deciduous trees or shrubs in the subtribe Pyrinae (formerly subfamily Maloideae; Potter et al., 2007) within the family Rosaceae. *S. aucuparia* L. (European mountain ash or rowan) [Figure 1a] grows into a small tree native to most of Europe. It is valued for its attractive shape, delicate leaves, autumn colours, and clusters of red or yellow berries. One particularly confusing name for rowans, used primarily in North America, is "mountain ash", which falsely implies that it is a species of ash (*Fraxinus*). The name arises from the superficial similarity in leaf shape of the two trees. In fact, rowan does not belong to the ash family, but is closely related to apples and hawthorns in the rose family. It is a perennial plant. The small white flowers are born in much-branched inflorescences and fruits are red. The bark is astringent and used in the treatment of diarrhea and as vaginal injection for leucorrhoea (Grieve, 1984; Chiej, 1984). Rowanberry (*S. aucuparia* L.) is a common, yellowish, wild berry that grows in the North of Europe. Berries have been described as an important source of flavonoids, particularly flavonols. Fruits of several *Sorbus* species (*S. aucuparia*, *S. domestica* and *S. torminalis*) are used as food ingredients (Berger, 1952; Hukkanen et al., 2006) and also as traditional diuretic, antiinflammatory, antidiarrhoeal (dried fruits), vasodilatory and vitamin agents with high antioxidant activity (Kähkönen et al., 1999; Kähkönen et al., 2001). Rowanberry is a good source of vitamin C (490 mg kg^{-1}) (Häkkinen et al., 1999). The seeds contain cyanogenic glycosides (amygdalin) which, after release of the glycosyl moiety, produce the extremely toxic prussic acid. In small quantities this acts as a stimulant to the respiratory system but in larger doses it can cause respiratory failure and death. It is therefore best to remove the seeds when using the fruit medicinally or as a food (Chevallier, 1996). Parasorbic acid, a characteristic constituent of the berries, is toxic. It is an unsaturated lactone [Figure 1b] that on hydrolysis gives a 2,4-hexadienoic acid, sorbic acid. The parasorbic acid occurs in the fruits in the form of glycoside, parasorboside. Whereas sorbic acid is a non-toxic substance and a permitted preservative in foodstuffs, parasorbic acid has local irritant properties that can lead to salivation, vomiting and gastroenteritis (Frohne and Pfänder, 1982, translated to English, 1984).

Like the fruits, the inflorescences of *S. aucuparia* are recommended in traditional medicine for similar disorders as diuretic and anti-inflammatory agents. There is a direct correlation

between the antioxidant capacity of *Sorbus* extracts and high content of phenolic compounds (Kähkönen et al., 1999).

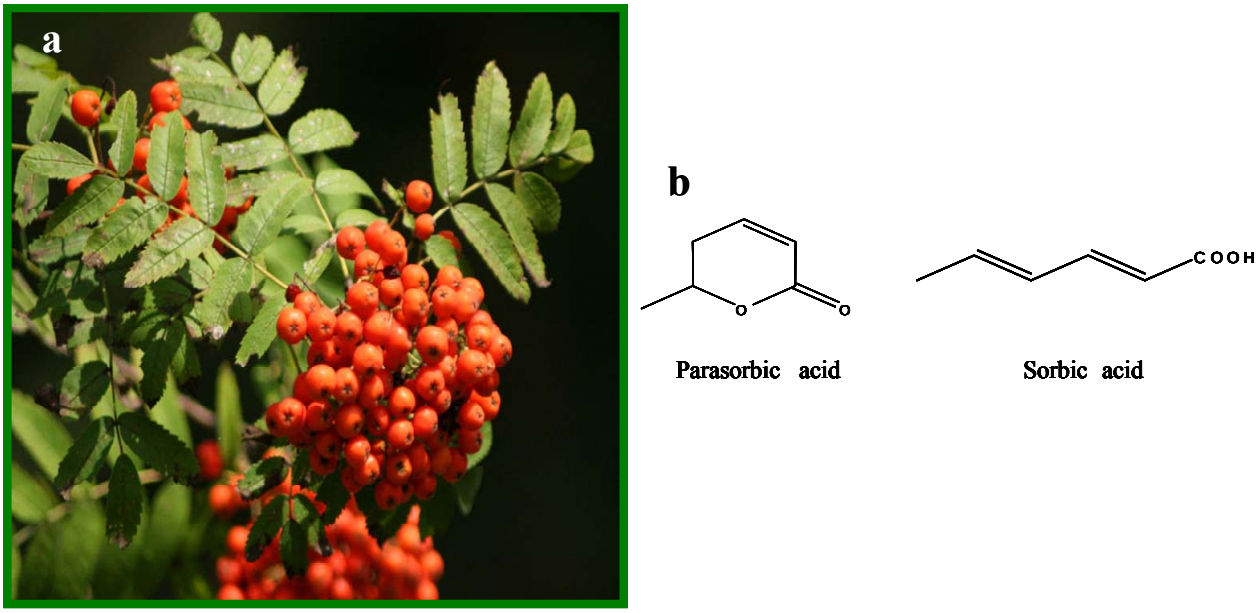


Figure 1. (a) *Sorbus aucuparia,* (b) characteristic active constituents.

2 Plant secondary metabolites

Plant secondary metabolite is a generic term used for more than 200,000 different substances. The plants form secondary metabolites e.g. for protection against pests, as colouring, scent, or attractants to pollinators and as the plant's own hormones (Wu and Chappell, 2008). It used to be believed that secondary metabolites were irrelevant for the human diet. However, secondary metabolites carry out a number of protective functions in the human body. These metabolites can boost the immune system, protect the body from free radicals, and kill pathogenic germs and much more (Makkar et al., 2007). In contrast to the primary metabolites (carbohydrates, fats, proteins, vitamins and mineral nutrients) secondary metabolites present in smaller amount and do not have nutrient characteristics for human beings. They have an effect on humans and health (Makkar et al., 2007). However many of these effects are still unknown. The exact requirement of the individual substances is likewise unknown. A diet which is rich in plant foods contains a variety of secondary metabolites and contributes to protecting the body against cancer and cardiovascular illnesses. Secondary metabolites and their effects are currently being intensively researched (Makkar et al., 2007). These metabolites are broadly classified into three major classes namely phenylpropanoids, terpenoids and alkaloids, mainly on the basis of starter molecules from which they are derived

(Croteau et al., 2000). The precursors are either primary metabolites or intermediate secondary compounds (De Luca and St. Pierre, 2000). Diversification of secondary metabolism mainly originated from molecular modifications such as oxidation, acylation, glycosylation and methylation.

2.1 Phenylpropanoids

Phenylpropanoids are so named because of the basic structure of a three-carbon side chain on an aromatic ring, which is derived from L-phenylalanine (L-Phe). All phenylpropanoids are derived from cinnamic acid, which is formed from phenylalanine by the action of phenylalanine ammonia-lyase (PAL). Severa1 simple phenylpropanoids (with the basic C_6-C_3 carbon skeleton of phenylalanine) are produced from cinnamate *via* a series of hydroxylation, methylation, and dehydration reactions; these include *p*-coumaric, caffeic, ferulic, and sinapic acids and simple coumarins. The free acids rarely accumulate to high levels inside plant cells; instead, they are usually conjugated to either sugars (e.g., salicylate-glucose conjugates), or cell wall carbohydrates (e.g., ferulate esters), and organic acids (e.g., sinapate esters, chlorogenic acid). Salicylic, benzoic, and *p*-hydroxybenzoic acids, although not strictly phenylpropanoids themselves because of the lack of the three carbons side chain, originate from the phenylpropanoids cinnamate and *p*-coumarate (Schnitzler et al., 1992; Yalpani et al., 1993). A large number of stress-induced phenylpropanoids [Figure 2] are derived from the C_{15} flavonoid skeleton, which is synthesized *via* chalcone synthase (CHS) (I.3.2).

Many phenylpropanoid compounds are induced in response to wounding or feeding by herbivores. Increased levels of coumestrol and coumarin are toxic to potential herbivores, causing estrogenic and anticoagulant effects, and psoralens can cause photo-induced blistering (Smith, 1982).

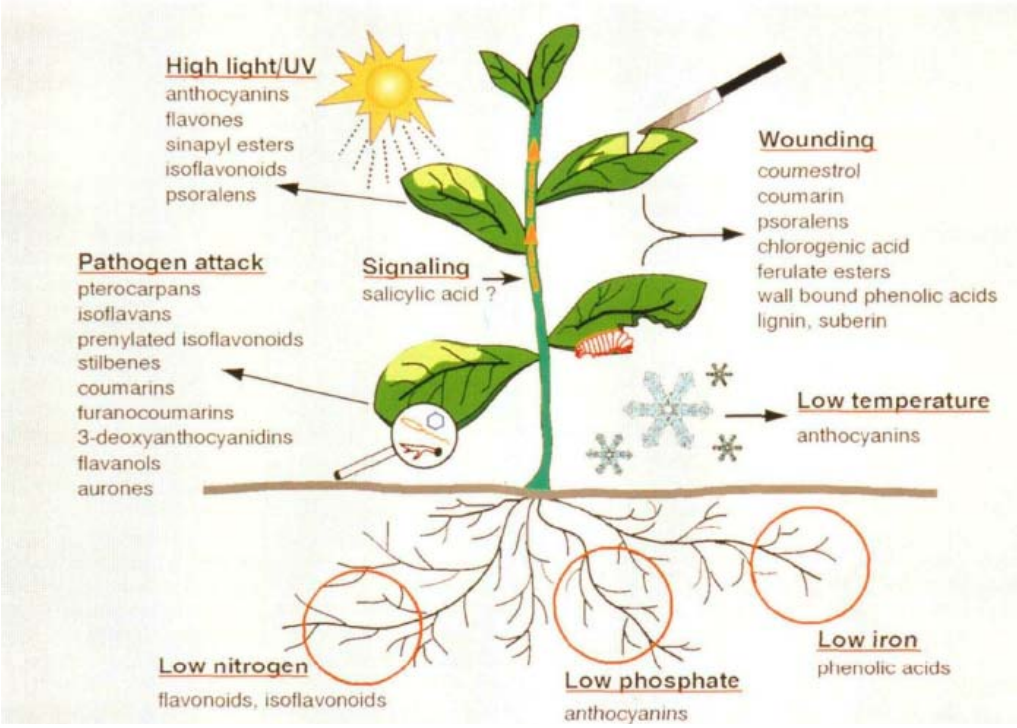


Figure 2. Examples of stress-induced phenylpropanoids (Dixon and Pavia, 1995).

2.1.1 Lignin

Lignin is a three-dimensional amorphous polymer consisting of methoxylated and hydroxylated phenylpropane units. It is essential to the life of vascular plants. The phenylpropane unit is a benzene ring with a tail of three carbons. In their natural unprocessed form, lignins are so complex that none of them has ever been completely described. They have huge molecular weights. Lignin is found in the cell walls of all vascular plants.

Hydroxycinnamoyl-CoA ester, a reaction product of the *p*-coumarate-CoA ligase-catalyzed reaction (I.3.1.2), further enters into the lignin biosynthetic pathway, leading to the formation of three types of phenylpropane units, also referred to as monolignols, (coniferyl-, coumaryl- and sinapyl-alcohol), and finally polymeric lignin by a multi-enzyme pathway [Figure 3] (Sarkanen and Ludwig, 1971; Lewis, 1999).

Since this polymerization process produces a polydisperse polymer with no extended sequences of regularly repeating units, its composition is generally characterized by the relative abundance of *p*-coumaryl, guaiacyl, and syringyl units (Lewis and Yamamoto, 1990; Bernards and Lewis, 1992).

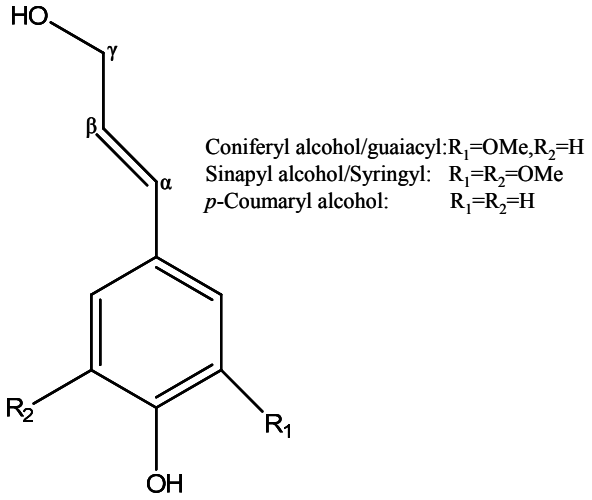


Figure 3. The three building blocks of lignin.

2.1.2 Flavonoids

Flavonoids are a diverse group of chemicals found in all plants. More than 4500 varieties of flavonoids have been identified, many of which are responsible for the attractive colors of flowers, fruits, and leaves (de Groot and Rauen, 1998). Flavonoids can be divided into various classes on the basis of their molecular structure (Rice-Evans et al., 1996; Harborne, 1988). The four main groups of flavonoids are flavones, flavanones, catechins and anthocyanins [Figure 4].

Starting from *p*-coumaroyl-CoA, the pathway also branches to synthesize flavonoid (C_6-C_3-C_6) and stilbene (C_6-C_2-C_6). The C_{15} flavonoid skeleton is formed by sequential decarboxylative addition of three molecules of malonyl-CoA to one molecule of *p*-coumaroyl-CoA, which is catalyzed by a type III polyketide synthase known as chalcone synthase (CHS) (Ferrer et al., 1999).

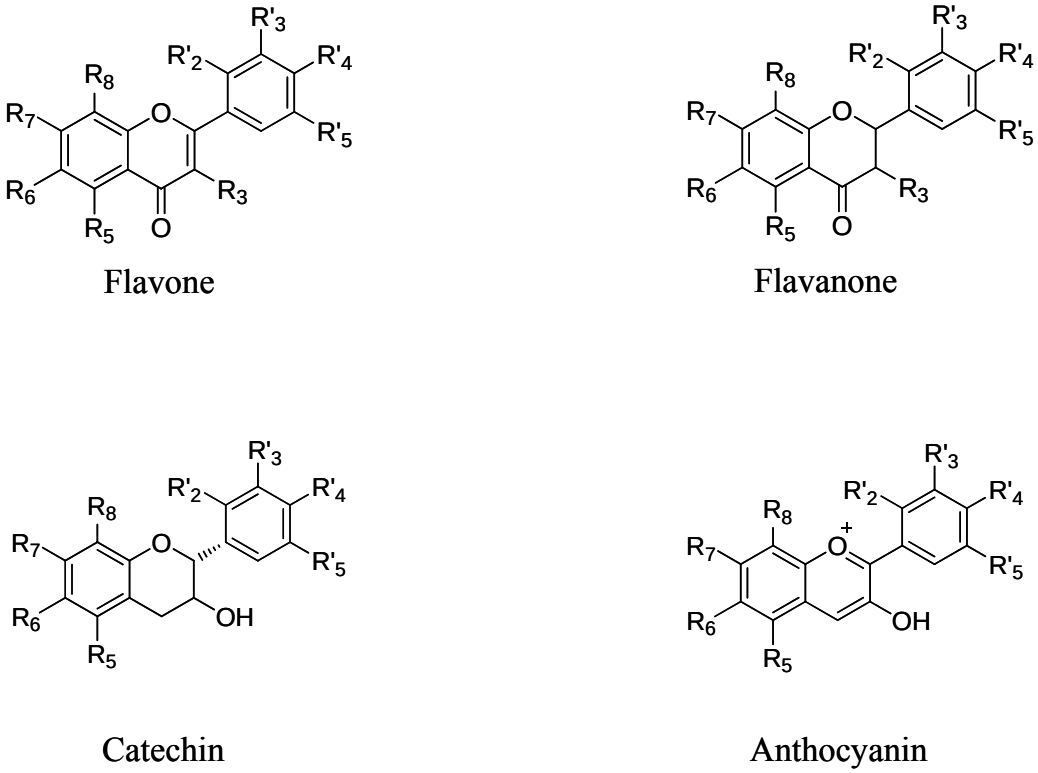


Figure 4. Chemical structures of the major flavonoid groups.

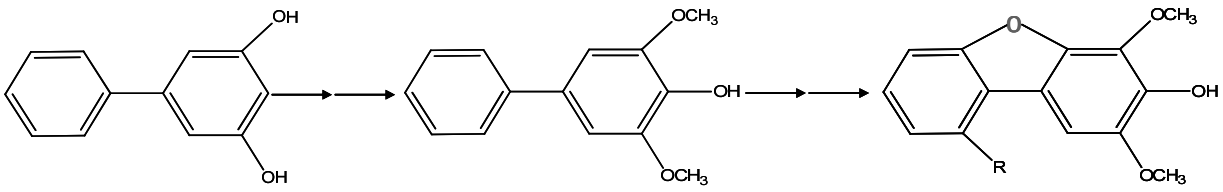
2.2 Phytoalexins

Phytoalexins are antimicrobial compounds of low molecular weight that are both synthesized by and accumulated in plants after exposure to microorganisms. Considerable evidence supports the view that accumulation of phytoalexins at the site of attempted infection is one mechanism by which a plant resists disease. One of the best evidence to date that phytoalexins are effective in providing plants with disease resistance has been obtained by transferring foreign phytoalexin expression from one plant to another. Stilbene phytoalexins, which include resveratrol, require only a single enzyme, stilbene synthase (STS), to link the two universally available precursors, malonyl-CoA and *p*-coumaroyl-CoA, for their synthesis.

Most phytoalexins commonly have a significant degree of lipophilicity. This may be important so that they can interact with the germination of the fungal spores in the aqueous environment of the leaf surface. Hardly has any of the many phytoalexins reported a glycosidic attachment (Harborne, 1999).

- **Phytoalexin variation in the Rosaceae**

Considering its large size (100 genera, 3000 spp) and its economic importance (many fruit crops), relatively little is known of the disease resistance mechanisms present in the family Rosaceae. The phytoalexins of the Pyrinae (formerly Maloideae), a subtribe of the Rosaceae, are biphenyls and dibenzofurans. It was reported that *Malus* species are always biphenyl producers, while *Pyrus* sp. produce exclusively dibenzofurans (Kokubun and Harborne, 1995) [Figure 5]. The best-known biphenyl phytoalexin is aucuparin, which was first isolated from *S. aucuparia* (Erdtman et al., 1963). Since then, biphenyls and dibenzofurans have been detected as phytoalexins following fungal infection in the sapwood of a number of species of the Maloideae, except for *S. aucuparia*, *Eriobotrya japonica*, and *Photinia glabra*, which formed the defence compounds in leaves (Miyakodo et al., 1985; Widyastuti et al., 1992; Kokubun and Harborne, 1994, 1995; Hrazdina, 2003). By contrast, *in vitro* cultures produced both biphenyls and dibenzofurans in response to elicitation. The major phytoalexins isolated from elicitor-treated cell cultures of *S. aucuparia* and a scab-resistant cultivar of *M. domestica* were aucuparin and the corresponding dibenzofuran derivative, eriobofuran [Figure 5] (Hrazdina et al., 1997; Borejsza-Wysocki et al., 1999; Liu et al., 2004; Hüttner et al., 2010). These two compounds were also detected in infected leaves of *Eriobotrya japonica* (Miyakodo et al., 1985).



3,5-dihydroxybiphenyl — aucuparin — eriobofuran : R=H; malusfuran : R=O-glucose

Figure 5. Proposed biosynthesis of the biphenyl aucuparin and dibenzofuran derivatives (Beerhues and Liu, 2009).

3 Biosynthesis of related secondary metabolites

3.1 Benzoic acids (BAs)

In plants, benzoic acid and its derivatives are important building blocks in a wide spectrum of compounds varying from primary metabolites like aromatic cytokinins and salicylic acid to secondary products with pharmacological activities such as the anti-cancer agent taxol and the local anesthetic cocaine.

To defend against pathogens, among other mechanisms, plants have developed hypersensitive response (HR) and systemic acquired resistance (SAR). HR leads to the formation of local necrotic lesions, restricting spreading of the pathogen. The primary infection that triggers HR may also result in the development of SAR, a long-lasting resistance to secondary infection, developed in the whole plant (Ross, 1961; Ryals et al., 1996). Both these responses are associated with production of salicylic acid (SA) adjacent to the infection site and in systemic parts of the plant (Enyedi et al., 1992; Me´traux et al., 1990; Yalpani et al., 1993; Klessig and Malamy, 1994).

Methyl salicylate may serve as an airborne signal activating defense-related genes in neighbouring plants and in the healthy tissue of the infected plant (Shulaev et al., 1997). 2,3-Dihydroxybenzoic acid and 4-hydroxybenzoic acid were reported to accumulate in elicitor-treated cell cultures (Kauss et al., 1993; Moreno et al., 1994; Sircar and Mitra, 2008, 2009). Shikonin arises from 4-hydroxybenzoic acid, and gallic acid is a component of gallotannins and ellagitannins (Dewick, 1997).

3.1.1 Biosynthesis of benzoic acids in plants

Plant BAs have been reported to originate either from L-Phe or directly from a shikimate-derived product such as isochorismate, in which the carboxyl carbon of shikimate is retained in the BA [Figure 6]. By contrast, the carboxyl carbon of BAs that are synthesized from Phe originates from the β-carbon of the Phe side chain.

Benzoic acid has also been reported to originate directly from shikimate *via* chorismic acid (Mustafa and Verpoorte, 2005; Wildermuth, 2006). Although direct origin of benzoic acid from shikimate/isochorismate is common in bacteria (Serino et al., 1995), only few reports are available in plants. In plants, direct origin of benzoate from shikimate was observed for gallic acid formation in *Rhus typina* (Werner et al., 1997) and 3-hydroxybenzoic acid formation in *Swertia chirata* root cultures (Wang et al., 2001, 2003) and *Centaurium erythraea* cell

cultures (Abd El-Mawla et al., 2001). Further, 2,3-dihydroxybenzoic acid accumulating in elicited-treated cell cultures of *Catharanthus roseus* and salicylic acid formation in *Arabidopsis thaliana* (Moreno et al., 1994; Wildermuth et al., 2001) have been postulated to originate from isochorismate.

Predominant early studies supported BA synthesis from Phe, and thus most work has focused on Phe-derived pathways. Few of the plant genes that encode the enzymes involved in BA biosynthesis have been cloned; thus, much work is needed to validate, define, and refine the biosynthetic pathways for BAs (Wildermuth, 2006). In the field of plant secondary metabolism, three routes from Phe to benzoic acid have been reported [Figures 6, 7]. The first one is a CoA-dependent and *ß*-oxidative-type pathway including four CoA-ester intermediates; the second route is a CoA-independent and non *ß*-oxidative pathway *via* benzaldehyde as a key intermediate; the third route is CoA-dependent yet non-*ß*-oxidative including cinnamoyl-CoA hydratase/lyase activity (HCHL) (French et al., 1976; Yazaki et al.1991; Schnitzler et al., 1992).

The CoA-dependent and *ß*-oxidative pathway involves activation of cinnamic acid, hydration of the CoA ester, oxidation of the hydroxyl group to a ketone, and cleavage of the *ß*-keto thioester *via* a reverse Claisen reaction to yield benzoic acid. This reaction mechanism is similar to the *ß*-oxidation of fatty acids (Zenk, 1965). This sequence appears to operate in elicited cell cultures of *Lithospermum erythrorhizon*, which accumulate shikonin, the biosynthesis of which proceeds *via* 4-hydroxybenzoic acid (Löscher and Heide, 1994). Benzoic acid biosynthesis in cucumber and *Nicotiana* species (Ribnicky et al., 1998; Jarvis et al., 2000) follows a similar chain-shortening mechanism using cinnamic acid as precursor.

In petunia, BA is a key intermediate in volatile benzenoid production. It serves as an intermediate precursor for benzoic acid/salicylic acid methyltransferase (PhBSMT) leading to the production of methylbenzoate (MeBA), the most abundant volatile emitted by petunia flowers (Negre et al., 2003). Recently, petunia 3-ketoacyl-CoA thiolase (PhKAT1), which plays an important role in the *ß*-oxidative pathway leading to production of benzoyl-CoA from 3-oxo-3-phenylpropionyl-CoA (benzoylacetyl-CoA), has been identified and characterized (Van Moerkercke et al., 2009).

The CoA-independent and non-*ß*-oxidative route involves hydration of the free acid, chain degradation *via* reverse aldol reaction, and oxidation of the intermediate aldehyde. Conversion of aromatic free acid, like *p*-coumaric acid, to 4-hydroxybenzaldehyde was found to be co-factor-independent and catalyzed by 4-hydroxybenzaldehyde synthase (Podstolski et al., 2002).

This pathway has been postulated to occur in cell cultures of *Daucus carota*, which incorporate 4-hydroxybenzoic acid into the cell walls upon elicitation (Schnitzler et al., 1992; Sircar, 2009). Potato tubers apparently use the same pathway to form benzoic acid (French et al., 1976). An analogous enzymatic reaction was observed in embryo cultures of *Vanilla planifolia* (Podstolski et al., 2002). However, in case of *V. planifolia*, the intermediate 4-hydroxybenzaldehyde gets converted into vanillin instead of *p*-hydroxybenzoic acid. Recently, a similar chain-shortening mechanism was observed in Tobacco Mosaic Virus (TMV)-infected *Nicotiana tabacum*, where 2-coumaric acid was converted into salicylaldehyde (Malinowski et al., 2007) by a salicylic aldehyde synthase (SAS).

A combination of two mechanisms, CoA dependent and non- *ß*-oxidative, has been detected in *Hypericum androsaemum* cell cultures [Figure 7] (Abd El-Mawla and Beerhues, 2002) and in transgenic hairy root cultures of *Datura stramonium* expressing a bacterial gene encoding hydroxycinnamoyl-CoA hydratase/lyase (Mitra et al., 2002). Starting from cinnamic acid, this route involves formation of cinnamoyl CoA, addition of a water molecule to the double bond of the cinnamoyl residue and subsequent cleavage of the putative intermediate 3-hydroxy-3-phenylpropionyl-CoA to yield benzaldehyde. This enzymatic reaction is catalyzed by a cinnamoyl-CoA hydratase/lyase (HCHL) type of enzyme. Finally, benzaldehyde is converted into benzoic acid in the presence of a NAD^+-dependent dehydrogenase.

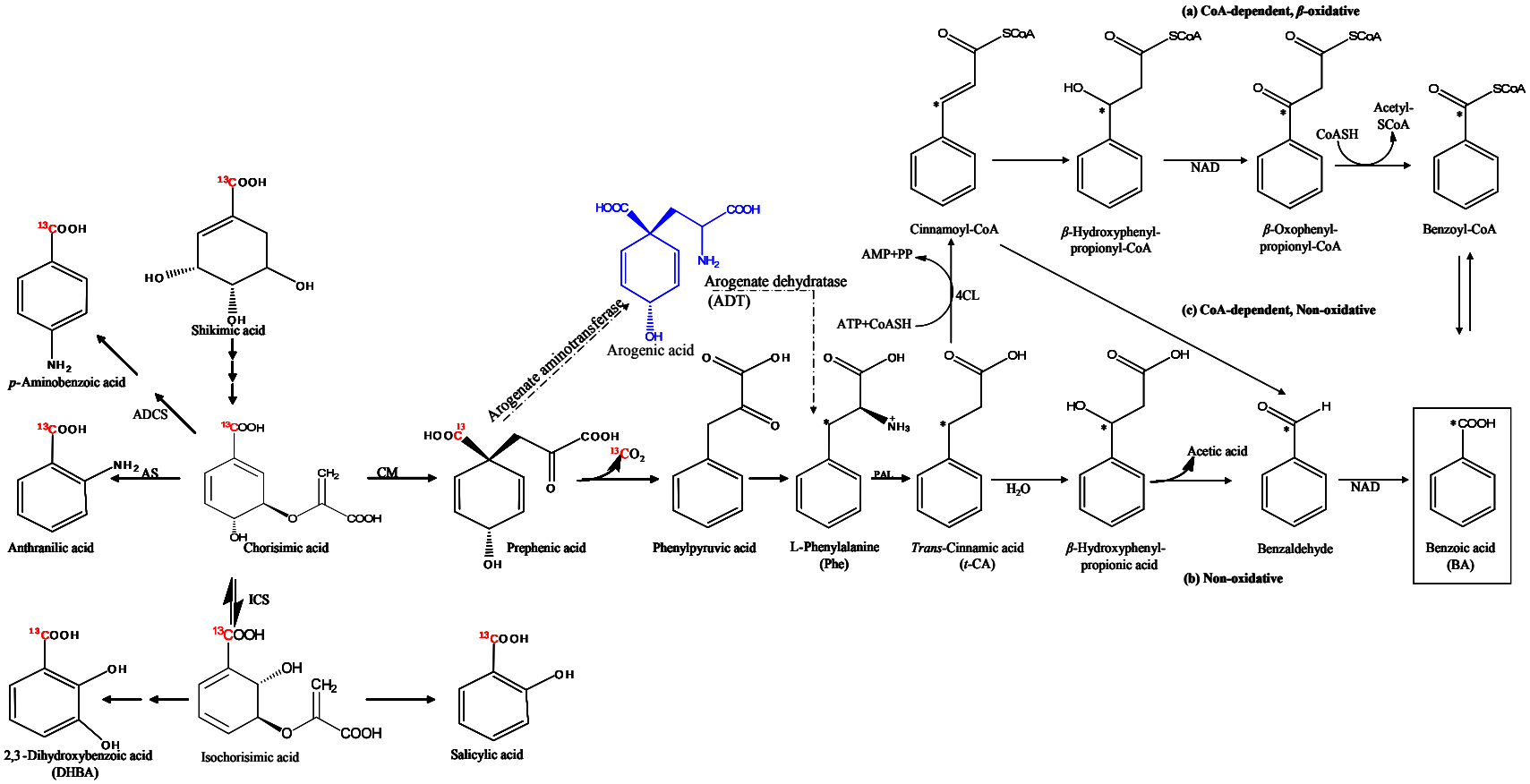


Figure 6. Biosynthesis of BAs in plants by the direct shikimate/chorismate pathway and *via* phenylalanine. The carboxyl carbon of shikimate is labeled (^{13}C) as is the β-carbon of phenylalanine (*). The plant enzymes involved in BA biosynthesis for which genes have been cloned are indicated, as are the chorismate-utilizing enzymes anthranilate synthase (AS) and aminodeoxychorismate synthase (ADCS). Pathways from *trans*-cinnamic acid alone are shown for simplicity; similar pathways from precursors in which either hydroxyl or methoxyl functionalities decorate the benzene ring (e.g. p-coumaric acid) have been reported. In addition, possible glucosylated precursors are not shown. C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CM, chorismate mutase; ICS, isochorismate synthase; PAL, phenylalanine ammonia lyase (Wildermuth 2006). Dashed arrows added by the author.

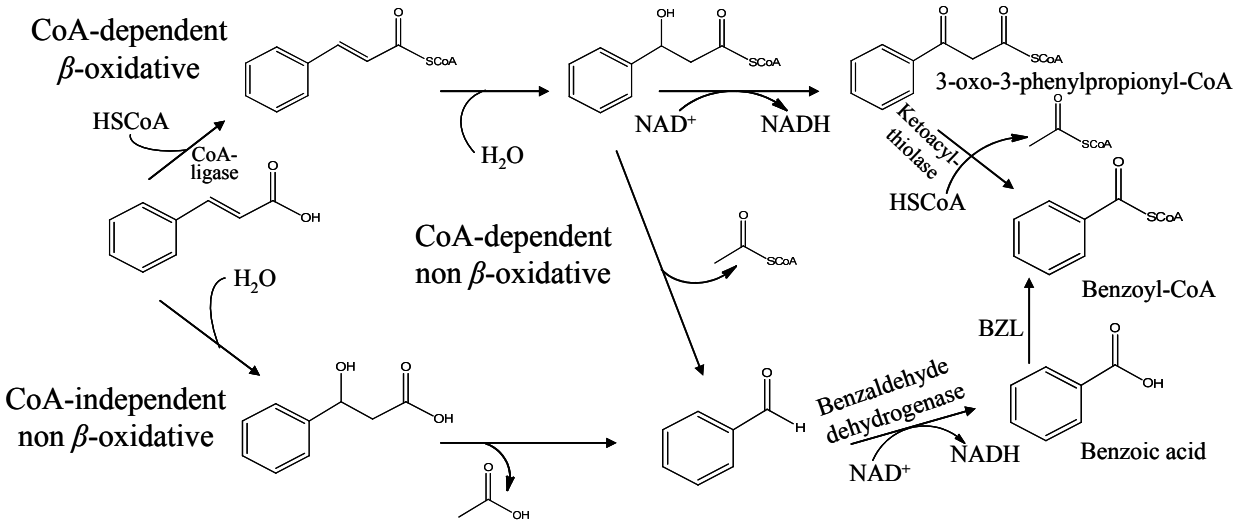


Figure 7. Postulated benzoic acid biosynthetic pathways (Abd El-Mawla and Beerhues, 2002).

3.1.2 Activation of benzoic acid

Incorporation of the benzoyl moiety into secondary metabolites occurs by the transfer of an activated BA intermediate, such as benzoyl-CoA, *via* an acyltransferase. During the first step of this reaction, an acyl-AMP intermediate is formed while pyrophosphate is released. In the second step the acyl-CoA ester is formed and AMP is released [Figure 8].

The enzyme *p*-coumarate-CoA ligase (4CL) is a well-characterized representative of this family, and its genes have been isolated from several plant sources (Lee and Douglas, 1996; Ehlting et al., 1999). 4CL is the last enzyme of the general phenylpropanoid metabolism and is required for the biosynthesis of a diverse array of plant natural phenylpropanoid products. It catalyzes the conversion of *p*-coumarate and other *p*-hydroxycinnamates to their corresponding CoA thiol esters. The general phenylpropanoid pathway connects primary metabolism with the biosynthesis of plant phenylpropanoid (I.2.1) natural products *via* the sequential actions of the enzymes phenylalanine ammonia-lyase, cinnamate 4-hydroxylase and 4CL [Figures 6].

The activated esters are used as precursors for the biosynthesis of important natural products formed *via* a variety of branched pathways. Significant examples include lignin, lignan, suberin, and flavonoids (I.2.1.1, I. 2.1.2) (Douglas, 1996; Hahlbrock and Scheel, 1989).

An enzyme that is functionally closely related to 4CL is cinnamate-CoA ligase which channels cinnamic acid from the general phenylpropanoid pathway into the benzoic acid biosynthetic route. Finally, benzoate-CoA ligase provides the starter substrate for BPS and BIS (benzophenone synthase and biphenyl synthase, respectively). Biphenyl synthase is the key enzyme of phytoalexin biosynthesis in Pyrinae (I.3.3).

Though benzoate-CoA ligase (BZL) had first been isolated and characterized from benzoate-degrading microorganisms (Lòpez et al., 2004; Schühle et al., 2003; Egland et al., 1995; Altenschmidt et al., 1991), the characterization of BZLs from plants is more recent. The enzymatic activity that catalyzes the formation of 3-hydroxybenzoyl-CoA (an intermediate in biosynthesis of xanthones) in crude extracts from cell cultures of *Centaurium erythraea* (Gentianaceae) has been reported (Barillas and Beerhues, 1997, 2000). In cell cultures of *Hypericum androsaemum* (Clusiaceae), a CoA ligase catalyzing the activation of benzoic acid has been detected (Abd El-Mawla and Beerhues, 2002). BZL from *Clarkia breweri* flowers has been purified and characterized (Beuerle and Pichersky, 2002). The enzyme exhibited, in addition to showing activity with BA, significant activity with other structurally related compounds, including anthranilic acid (50% relative activity compared with BA) and 3-hydroxybenzoic acid (25% relative activity). BZO1 (At1g65880) encodes an enzyme with benzoyl-CoA ligase activity required for the accumulation of benzoyloxyglucosinolates in *Arabidopsis* (Kliebenstein et al., 2007).

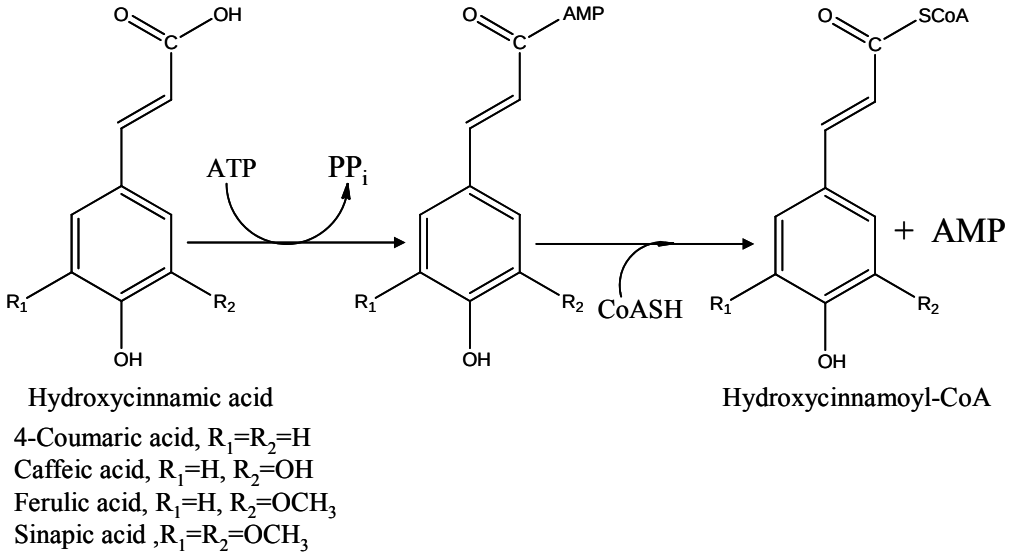


Figure 8. Reaction catalyzed by *p*-coumarate-CoA ligase.

3.2 Biosynthesis of flavonoids

In most plant families the initial product of CHS, the key enzyme in flavonoid biosynthesis, is flavonoid classes (flavones, flavanones, catechins, anthocyanins and 3-deoxyanthocyanidins, Holton and Cornish, 1995). In legumes, which possess chalcone reductase (CHR) in addition to CHS, a trihydroxychalcone may be formed (Welle and Grisebach, 1989). In a number of species, including pine, grapevine, and peanut, the condensation of *p*-coumaroyl-CoA or cinnamoyl-CoA with three malonyl-CoA molecules can also give rise to stilbenes by the action of stilbene synthase (STS) (Schröder et al., 1988). In legumes, isoflavone synthase (IFS) rearranges the flavonoid carbon skeleton, leading to the accumulation of a wide range of simple isoflavonoids, coumestans, pterocarpans, and isoflavans. Structural diversity among these derivatives is brought about by a variety of modifications, including regiospecific hydroxylation, glycosylation, acylation, prenylation, sulfation, and methylation.

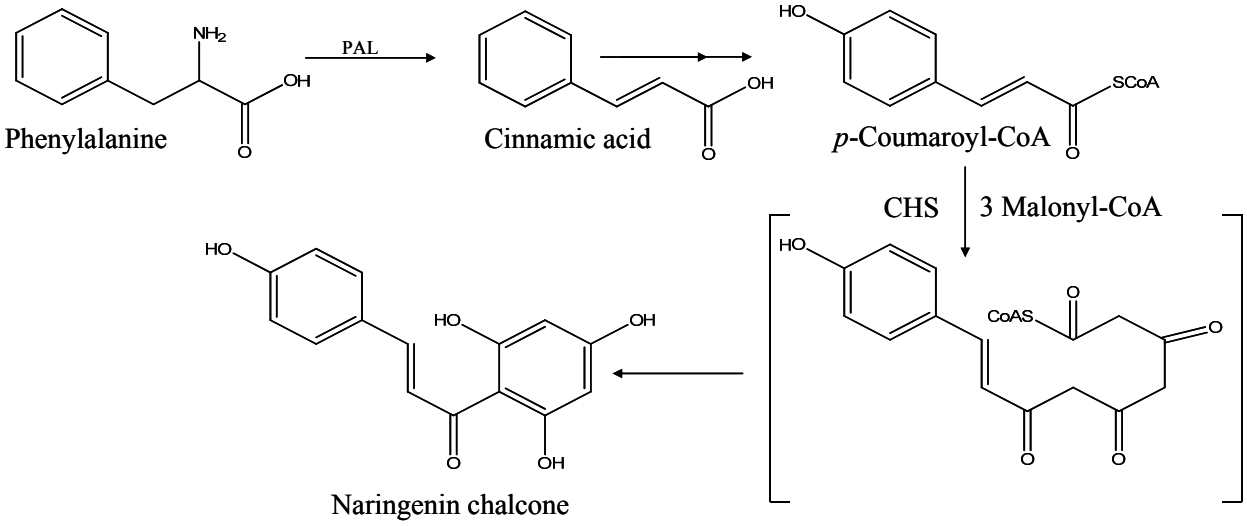


Figure 9. Flavonoid biosynthetic pathway.

3.3 Biosynthesis of polyketides other than flavonoids

The polyketide natural products are one of the largest and most diverse groups of secondary metabolites. Polyketides are produced by many different organisms, from protists and plants to bacteria. Many diverse polyketides are synthesized by bacteria and fungi and have antibiotic or mycotoxic properties (e.g., erythromycin, rifamycin and actinorhodin). Polyketide compounds are also synthesized by plants, in which they have diverse functions including roles in flower pigmentation, pathogen defense (phytoalexins), response to UV and visible light exposure, and symbiotic plant–pathogen interactions (Schröder et al., 1998; Winkel-Shirley, 2002). The diversity of biological activity of polyketides has made these secondary metabolites, and the polyketide synthases (PKSs) that synthesize them, an important focus of biopharmaceutical research. PKSs are classified according to their architectural configurations as type I, II and III (Fischbach and Walsh, 2006; Hopwood and Sherman, 1990; Staunton and Weissman, 2001). Type III PKSs are homodimeric enzymes of 40-45 kDa subunits. They are a group of enzymes that generate a diverse array of natural products by catalyzing the condensation of a starter CoA ester, such as acetyl-CoA, with extender CoA esters, such as malonyl-CoA (Austin and Noel 2003; Schröder, 1999). The homodimeric enzymes orchestrate a series of acyltransfer, decarboxylation, condensation, cyclization, and aromatization reactions at two functionally independent active sites. CHS from *Medicago sativa* is the biochemically and structurally best characterized type III PKS (Ferrer et al., 1999).

Among the starter substrates used, benzoyl-CoA is a rare starter molecule. It is utilized by microbial type I PKSs to form soraphen A, and type II to from enterocin and the wailupemycins (Moore, 2005). In plants, benzoyl-CoA is the starter unit for two type III PKSs, benzophenone synthase (BPS) and biphenyl synthase (BIS). Biphenyl synthase is the key enzyme of phytoalexin biosynthesis in Pyrinae and is a relatively recently detected type III PKS (Liu et al., 2004; 2007).

BIS and BPS form identical linear tetraketides [Figure 10]. However, while BPS cyclizes this intermediate *via* an intramolecular C6→C1 claisen condensation, BIS catalyzes an intramolecular C2→C7 aldol condensation and decarboxylative elimination of the terminal carboxyl group to give 3,5-dihydroxybiphenyl. When benzoyl-CoA is replaced with salicoyl-CoA as a starter substrate, BIS catalyzes a single decarboxylative condensation reaction with malonyl-CoA to form a diketide intermediate, which undergoes intramolecular cyclization by

nucleophilic attack of the phenol group on the CoA- or cysteine-tethered C-1 thioester, yielding 4-hydroxycoumarin after enolization (Liu et al., 2010).

cDNAs encoding BIS and BPS were cloned from cell cultures of *S. aucuparia* and *H. androsaemum*, respectively. The enzymes share 54% amino acid sequence identity over their approx. 400 amino acids (Liu et al., 2003, 2007; Beerhues et al., 2007; Beerhues and Liu, 2009). A point mutation in the active site cavity (Thr 135 Leu) transformed BPS into a functional phenylpyrone synthase (PPS) (Klundt et al., 2009). PPS is a promising biotechnological tool for manipulating benzoate-primed biosynthetic pathways, such as benzophenone and biphenyl biosyntheses in Clusiaceae and Pyrinae, respectively.

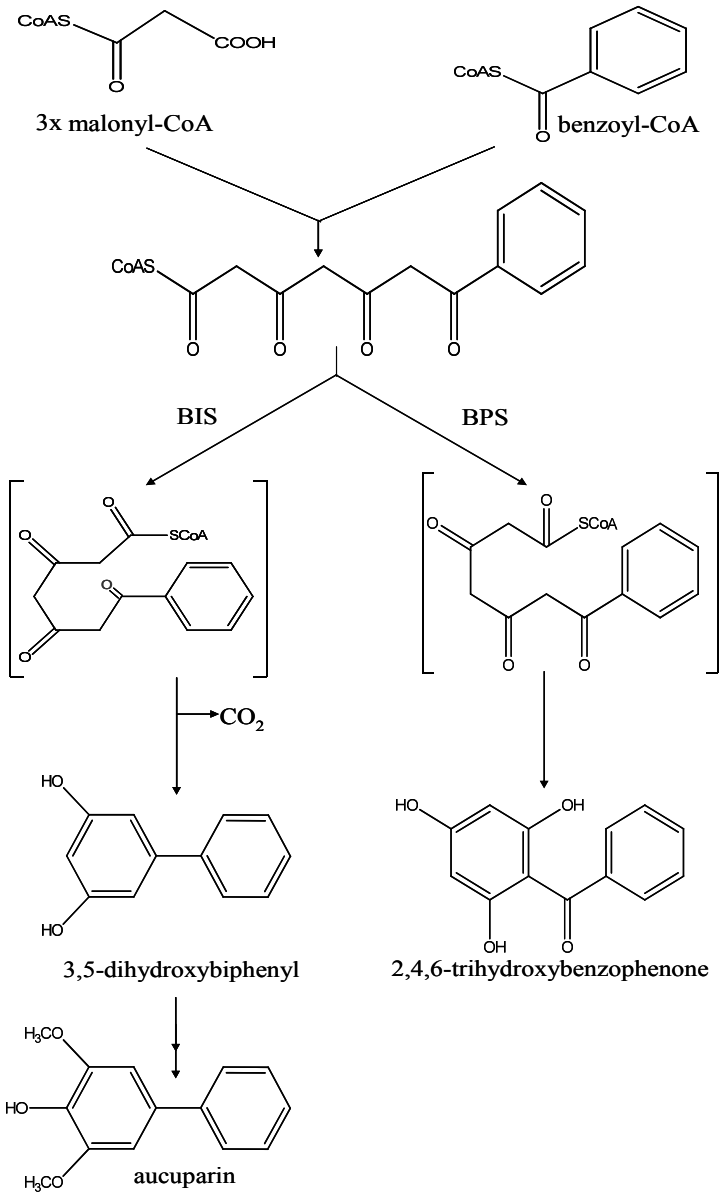


Figure 10. Reactions of biphenyl synthase (BIS) and benzophenone synthase (BPS).

4 Research Strategies and Objectives

Despite the wide distribution of benzoic acid-derived secondary metabolites, little is known about the biosynthesis and activation of benzoic acid. Therefore, this thesis aims at exploring the enzymatic route of benzoic acid formation in *Sorbus aucuparia* cell suspension cultures. This *in vitro* system was previously used to isolate cDNAs encoding BIS isoforms which are phytoalexin-forming enzymes in Pyrinae (Liu et al., 2007, 2010).

The major objectives of this work were:

- To detect and characterize benzaldehyde dehydrogenase (BD) in chitosan-treated cell cultures of *S. aucuparia.*

- To study the coordinated induction between BD and BIS to collect evidence for BD involvement in biphenyl biosynthesis

- To detect a cDNA coding for BD.

- To use a homology-based approach for isolating CoA-ligase cDNAs related to secondary metabolism in elicitor-treated *S. aucuparia* cell cultures.

- To functionally express the recombinant proteins and to study their contribution to benzoyl-CoA formation.

Detection and characterization of these enzymes provide the first insight into benzoic acid metabolism in the economically valuable taxon, Pyrinae (formerly Maloideae).

II. Materials and Methods

1 Materials

1.1 Plant material

Tissue cultures of *Sorbus aucuparia* were established and grown as described previously (II.2.1; Liu et al., 2004).

Figure 11. Cell suspension cultures of *Sorbus aucuparia* on the day of subculturing (left) and after 14 days of growth (right).

1.2 Chemicals

For preparation of culture media, buffers and solutions deionized water was used, which was obtained from an Arium 611 VF water purification system (Sartorius, Germany). All solutions were sterilized by autoclaving for 20 minutes at 121°C and a pressure of 2 bars. All antibiotics and other thermolabile substances were sterile-filtered, then added after autoclaving. Commonly used chemicals and glass wares, unless otherwise specified, were mostly purchased from Sigma-Aldrich, Fluka, Roth, Bio-Rad, Riedel-deHaën, Serva, or Fisher Scientific.

Chemical	Supplier
Phytohormones for plant cultures	
NAA	Fluka
2,4-D	Fluka
Elicitor for plant cultures	
Chitosan	Roth
Reagents for bacterial culture medium	
Agar-Agar	Roth
Peptone	Roth
Yeast extract	Roth
Sucrose	Fluka
KH_2PO_4	Roth
K_2HPO_4	Roth
$MgSO_4 .7H_2O$	Fluka
Glycerin	Roth
Antibiotics	
Ampicillin	Roth
Chloramphenicol	Fluka
Reagents for biochemistry and molecular biology	
DMSO	Fluka
IPTG	Sigma-Aldrich
X-Gal	Sigma-Aldrich
dNTPs	Fermentas
Glutathione (Reduced form)	Sigma
TRIS-HCl	Roth
Imidazole	Roth
Reagents used for crude protein extraction from cell culture	
Polyclar®AT	Serva
Seasand	Roth
HEPES	Sigma-Aldrich
Dithiothreitol (DTT)	Diagnostic Chemicals Limited (dcl)
Chemicals for radioisotopic assay	
Benzoic acid [7-^{14}C]	Hartmann Analytik
Indicator cocktail Lumasafe ^{TM} Plus	LUMAC LSC
Reagents for GC-MS derivatization	
N-methyl-N-(tri-methylsilyl)-trifluoroacetamide (MSTFA)	ABCR
Methoxyamine hydrochloride	Sigma-Aldrich
Substrates, cofactors and authentics for enzyme assays	
ATP	Roth
Benzoic acid and its derivatives	Sigma-Aldrich /Fluka
NAD^+, $NADP^+$, FAD, FMN	Sigma / GERBU Biotechnik
$FeCl_2$,$MgCl_2$,$CaCl_2$,$MnCl_2$,$CoCl_2$,$ZnCl_2$,$CuCl_2$	Roth/Fluka/Merck/Riedel-de Haën

Coenzyme A	Sigma-Aldrich
Benzaldehyde and its derivatives	Sigma-Aldrich
Cinnamic acid and its derivatives	Sigma-Aldrich/Fluka
Benzoyl-CoA	Sigma-Aldrich
Malonyl-CoA lithium salt	Sigma-Aldrich
3,5-Dihydroxybiphenyl	Synthesized in our group (Liu et al., 2004)

Reagents used to stop enzymatic reaction

Trichloroacetic acid	Fluka
Glacial acetic acid	Roth

Stationary phases used for protein affinity purification, desalting and concentration

Protino® Glutathione Agarose 4B	Macherey-Nagel
Ni-NTA agarose	Qiagen
PD_{10} –Cartridge Sepharose G-25 columns	GE Healthcare
Vivaspin 6 Centrifugal Concentrators	Satorius Stedim Biotech

Reagents for gel electrophoresis

peqGOLD Universal Agarose	Peqlab
Ethidium bromide	Roth
Urea	Bio-Rad
TEMED	Bio-Rad
Acrylamide / Bisacrylamide 30%	Bio-Rad
Ammonium persulfate	Roth
EDTA	Sigma-Aldrich
Formamide	Sigma-Aldrich
SDS	Roth
β-mercaptoethanol	Fluka
Bromophenol blue	Sigma
Coomassie-blue R 250 and G 250	Merck

Solvents for HPLC

Methanol	Fisher-Scientific
Orthophosphoric acid	Roth
Trifluoroacetic acid	Roth
Acetonitrile	Fisher-Scientific

Ladders

Prestained protein ladder (10-170 kDa)	Fermentas
Gene Ruler DNA Ladder Mix	Fermentas

1.3 Nutrient media

1.3.1 Medium used for growing cell suspension culture

LS medium (Linsmaier and Skoog, 1995)	Stock solution ingredients			Supplier	For 1 liter medium
Macro elements	KNO_3	19.0	g	Sigma-Aldrich	100 ml
	NH_4NO_3	16.5	g		
	$CaCl_2 \cdot 2H_2O$	4.4	g		
	$MgSO_4 \cdot 7H_2O$	3.7	g		
	KH_2PO_4	1.7	g		
	$Na_2EDTA \cdot 2 H_2O$	0.41	g		
	$FeSO_4 \cdot 7 H_2O$	0.28	g		
	H_2O ad	1	L		
Micro elements	$MnSO_4 \cdot H_2O$	16.9	g	Sigma-Aldrich	1 ml
	$ZnSO_4 \cdot 7 H_2O$	10.6	g		
	KI	0.83	g		
	H_3BO_3	6.2	g		
	$Na_2MoO_4 \cdot 2 H_2O$	0.25	g		
	$CuSO_4 \cdot 5 H_2O$	0.025	g		
	$CoCl_2 \cdot 6 H_2O$	0.025	g		
	H_2O ad	1	L		
Vitamins	Thiamine-HCl	4	mg	Serva Sigma	10 ml
	Myo-Inositol	1000	mg		
	H_2O ad	100	ml		
Hormones	2,4-D	1	mg	Fluka	220 µl
	Ethanol	1	ml		
	NAA	1	mg	Fluka	186 µl
	Ethanol	1	ml		
Sucrose	30g			Fluka	
pH	6				
Solid LS (for callus cultures)	0.8% Agar			Roth	

1.3.2 Bacterial culture medium and reagents

Medium	Components	
Luria - Bertani LB medium (Sambrook and MacCallum, 2001) For solid medium	Bacto-pepton Yeast extract NaCl Agar	10 g/l 5 g/l 10 g/l 1.5%
SOC-Medium	Tryptone/Peptone from Casein Yeast extract 1 M NaCl 1 M KCl Autoclave, then add: Sterile filtered solution of 2 M Mg^{2+} 2 M glucose	20 g/l 5 g/l 10 ml/l 2.5 ml/l 10 ml/l 10 ml/l
For -80°C storage of bacterial cultures	-Bacterial culture in LB medium -Autoclaved solution of : Glycerin (Roth):LB (2:8)	250 µl 750 µl
Antibiotic	-Ampicillin 150 µg/ml medium -Chloramphenicol 60 µg/ml medium (prepared in pure ethanol)	
For induction of bacterial culture	IPTG Final concentration used for induction (Sterile-filtered and freshly prepared)	0.5 M 1mM
For blue/white screening *X*-Gal	5-Bromo-4-chloro-3-indolyl-*β*-D- galactopyranosid N,N`-Dimethylformamid (Stored at -20°C) Use 40 µl per petridish	 40 mg 1 ml

1.4 Buffers and solutions

1.4.1 Buffers and solutions for gel electrophoresis

DNA-electrophoresis		
50x TAE buffer	Tris-HCl	2 M
	EDTA	0.05 M
	pH 8 with glacial acetic acid	

Agarose-gels were prepared at concentrations of 0.8-2% agarose in electrophoresis buffer (1x TAE)

For sequencing gel		
TBE buffer	Tris-HCl	90 mM
	Boric acid	90 mM
	EDTA	2 mM
Loading buffer	25 mM EDTA	1 ml
	Formamide	5 ml
	Dextran blue	0.1 mg
Gel	Urea	9 g
	10x TBE buffer	3 ml
	dH_2O	11.5 ml
	30% Acrylamide/Bis. (Bio-Rad)	3.75 ml
	TEMED (Roth)	10 µl
	10% (w/v) APS	200 µl

Protein electrophoresis (SDS-PAGE)		
Stacking gel (5%)	dH_2O	3.4 ml
	0.5 M Tris-HCl (pH 6.8)	0.63 ml
	30% Acrylamide/Bis.	0.83 ml
	10% (w/v) SDS	0.05 ml
	10% (w/v) APS	0.05 ml
	TEMED	5 µl
Separating gel (12%)	dH_2O	3.3 ml
	1.5 M Tris-HCl (pH 8.8)	2.5 ml
	30% Acrylamide/Bis.	4 ml
	10% (w/v) SDS	0.01 ml
	10% (w/v) APS	0.01 ml
	TEMED	4 µl

TEMED is used with APS to catalyze the polymerization of acrylamide when making polyacrylamide gels, used in gel electrophoresis for the separation of proteins or nucleic acids.

Protein loading buffer	dH_2O	2.7 ml
	0.5 M Tris-HCl (pH 6.8)	1 ml
	Glycerin	2 ml
	10% (w/v) SDS	3.3 ml
	β-mercaptoethanol	0.5 ml
	0.5% (w/v) Bromophenol blue	0.5 ml
10x Electrophoretic buffer	Tris-HCl	15 g
	Glycin	72 g
	Na-SDS	5 g
	dH_2O	ad 500 ml
Staining solution	Coomassie-blue R 250	25 ml
	Methanol	100 ml
	Acetic acid	20 ml
	Water	ad 200 ml
Coomassie-blue stock solution	Coomassie-blue R 250	0.5 g
	Water	ad 50 ml
	Filtered before use	
Destaining solution	Methanol	30 ml
	Acetic acid	20 ml
	dH_2O	200 ml

1.4.2 Buffers for affinity purification of fusion protein

1.4.2.1 GST-fusion protein

Lysis buffer 0.1 M KH_2PO_4 pH 7-7.5	KH_2PO_4 dH_2O	1.36 g ad 100 ml
Binding buffer PBS buffer pH 7	NaCl KCl Na_2HPO_4 KH_2PO_4 dH_2O	4.09 g 0.10 g 0.89 g 0.12 g ad 500 ml
Elution buffer pH 8	50 mM Tris-HCl 10 mM Glutathione (reduced) dH_2O	0.6 g 0.3 g ad 100 ml

1.4.2.2 His_6-tagged fusion protein

Lysis buffer pH 8	NaH_2PO_4 NaCl 20 mM Imidazole dH_2O	3.4 g 0.9 g 0.7 g ad 500 ml
Washing buffer pH 8	NaH_2PO_4 NaCl 50 mM Imidazole dH_2O	3.4 g 0.04 g 1.7 g ad 500 ml
Elution buffer pH 8	NaH_2PO_4 NaCl 250 mM Imidazole dH_2O	3.4 g 8.8 g 8.5 g ad 500 ml

1.4.3 Buffer used for crude protein extraction from cell cultures

HEPES buffer was always used for crude protein extraction from plant cell suspension culture.

HEPES buffer	HEPES (100 mM) DTT (10 mM) Water	2.4 g 0.154 g ad 100 ml
Adjust pH to 8 by adding 1N NaOH. DTT should be added freshly.		

1.4.4 Buffers used for enzyme assays

Enzyme to be assayed	**Buffer**	
Benzaldehyde dehydrogenase (BD)	Tris-HCl (200 mM) dH_2O Adjust pH to 9.5 by conc. HCl	2.4 g ad 100 ml
p-coumarate-CoA ligase (4CL) and biphenyl synthase (BIS)	Tris-HCl (100 mM) dH_2O Adjust pH to 7.5 by conc. HCl	1.21 g ad 100 ml
BD and 4CL pH optimization at acidic range	KH_2PO_4 (100 mM) dH_2O pH adjusted by KOH	1.36 g ad 100 ml
BD and 4CL pH optimization at alkaline range	Tris-HCl buffer used	

1.4.5 Buffers used for plasmid isolation (mini prep)

Buffer 1	Tris-HCl	50 mM
	EDTA	10 mM
	RNase A	100 µg/ml
	Adjust to pH 8 with HCl	
	RNase A was freshly added prior to use	
Buffer 2	NaOH	0.2 M
	SDS	1% (w/v)
Buffer 3	Potassium acetate	2.55 M
	Adjust to pH 5.5 with glacial acetic acid	

1.4.6 Solution for protein estimation

Bradford-Reagent	Coomassie-Brilliant blue G250	100 mg
	Ethanol 96%	50 ml
	Phosphoric acid 85%	100 ml
	dH_2O	ad 1000 ml

Dissolve well Coomassie®-Brilliant G250 in ethanol, add orthophosphoric acid and make volume up to 1 L with water. Filter the solution through filter paper (Whatman No. 1) until no blue colour can be seen. Keep at 4°C in amber glass bottle.

1.4.7 Washing solution for regeneration of PD_{10} column

NaOH 0.16 M	Wash PD_{10} column with five column volumes of NaOH cleaning solution followed by five volumes of distilled water. At the end, washings pH should be neutral.

1.5 Materials used for molecular biology

1.5.1 Host cells and Cloning vector

1.5.1.1 Host cells (Competent *E. coli*)

E.coli	Purpose	Genotype
DH5α	This strain was used for the initial cloning of target DNA into pGEM-T easy vector.	F' φ80δlacZ9M15 end A1 hsdR17(rk-mk+)supE44thi-1 λ-gyrA96 relA1 9(lacZYA-argFV169) deoR
BL21(DE3)pLysS	Used for protein expression of target gene cloned in pGEX vector.	F^- *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm* (DE3) pLysS (Cam^R)

1.5.1.2 Vectors

Vectors	Characters/purpose	Supplier
A-Cloning vector		
pGEM-T easy	3kb T-overhanged vector with *lacZ* and ampicillin resistance genes for subcloning of *Taq* DNA polymerase-amplified PCR products.	Invitrogen
B-Expression vectors		
pRSET B	2,9 kb expression vector with N-terminal His_6-tag and ampicillin resistance gene	Invitrogen
pGEX-G	4,9 kb expression vector with N-terminal GST-tag and ampicillin resistance gene	Smith and Johnson, 1988; Görlach and Schmid, 1996

1.5.2 Primers

All primers were synthesized in HPSF (High Purity Salt Free) quality at MWG-Biotech AG (Ebersberg, Germany)

1.5.2.1 Reverse Sa4CL1 primers used to amplify Sa4CL1 5´-end

Name	Gene specific primer
Rev-P1	5`-CGGAGAGTCGACGCACATTAG-3`
Rev-P2	5`-GTGAACATCGTCACTTGATGACG-3`
Rev-P3	5`-CTTGTCGTAGTAGCACGCTAAG-3`
Rev-P4	5`-AAAGGAGGCTCCGAGGAAGG-3`

1.5.2.2 Forward Sa4CL-4 primers used to amplify 3`-end

F-P1	5`-ATGGCTATAGAAACTACCCAAAACGAC-3´
F-P2	5`-CTCAGAGTTGCCTTAGGGCTTAACAAG-3`
5`UTR-primer*	5`-CTATATGAAGATCCCTATGGGCC-3`

* A primer designed in the 5´- untranslated region (5´UTR). Together with 3´-RACE primers (II.1.5.2.5) it could amplify SaALDH (III.3)

1.5.2.3 Sa4CL1 over-expression primers

Sa.Exp.1-3`***Kpn*I**	5`-ATT **GGT ACC** TCA GTT TGG AAA CCC AG -3`
Sa.Exp.1-5` ***Nhe*I**	5`-ATT **GCT AGC** ATG GCT ATA GAA ACT ATC C-3`
Exp.4CL1-ATG	5`-ATG GCT ATA GAA ACT ATC CCA AAC GAC ATC-3`

Sa.Exp.1-3`***Kpn*I** & Sa.Exp.1-5` ***Nhe*I** are overexpression primers for pRSET B expression vector.
Sa.Exp.1-3`***Kpn*I** & Exp.4CL1-ATG are overexpression primers for pGEX expression vector.
The underlined sequence is spacer nucleotides (See II.3.1), the bold sequences are restriction sites.

1.5.2.4 Vector specific primers (for sequencing the DNA insert)

pGEM-T easy	T7	5`-TTG TAA TAC GAC TCA CTA TAG-3`
	Sp6	5`-GAT TTA GGT GAC ACT ATA GAA TAC-3`
pRSET B	Forward	5`-GAG ACC ACA ACG GTT TCC CTC-3`
	Reverse	5`-CTA GTT ATT GCT CAG CGG TGG-3`
pGEX-G	Forward	5`-ATA GCA TGG CCT TTG CAG G-3`
	Reverse	5´-GAG CTG CAT GTG TCA GAG G-3`

1.5.2.5 SMART II-RACE primers

SMART II	5`-AAG CAG TGG TAA CAA CGC AGA GTA CGC GGG-3`
3`CDS	5`-AAG CAG TGG TAA CAA CGC AGA GTA $C(T)_{30}$ $N_{-1}N$-3`
5`CDS	5`-$(T)_{25}$ $N_{-1}N$- 3`
RACE-Short	5`-CTA ATA CGA CTC ACT ATA AGG GC-3`
RACE-Nested	5`-AAG CAG TGG TAA CAA CGC AGA GT-3`
RACE-Long	5`-CTA ATA CGA CTC ACT ATA AGG GCA AGC AGT GGT AAC AAC GCA GAG T-3`

1.5.2.6 TATA Box degenerate primer (TD)

5`-GAT TCT AGA (CT)(CT)I CTA TA(AT) A(AT)A (GC)(AC)-3`

1.5.2.7 Primers used for expression analysis by RT-PCR

Transcript	Primer	Sequence	Reference
PAL (1070 bp)	forward	5`-ACA CTC CTC CAA GGC TAC TCC-3`	The present study
	reverse	5`-ATG GAC TGA CAA AGC GCA ACC-3`	
Sa4CL1 (898 bp)	forward	5`- TCC CAA CTT GTA TTA TAG CAC C -3`	Scharnhop, 2008
	reverse	5`- TTC GAT GAA AAA TAC TCG ATT TAT TC -3`	
Sa4CL2 (897 bp)	forward	5`- TCC GAA TTT GTA TTT CCA CAG T -3`	
	reverse	5`- TCT GTG AAG AAA ACC CGA C - 3`	
Sa4CL3 (896 bp)	forward	5`- CAA ACC TCT ACT TGA AGG AG -3`	
	reverse	5`- ATG GAC AAA GTG CAC CTT GT- 3`	
BIS3 (1313 bp)	forward *	5`-CAA CAA ATC CCT ACT CTA TAT CTG C-3`	Liu et al., 2010
	reverse **	5`-TGA AAC TCG ATC GGT TGC AGT AGC-3`	
BIS1 (1173 bp)	forward	5`-TGG CGC CTT TGG TTA AGA ATC A-3´	Liu et al., 2007 & 2010
	reverse	5`-TTA GCA TGG AAT AGA TTC ACT ACG CAG -3'	
SaALDH (548 bp)	forward	5`-GCT GCC TTC GAC CAT GGC TCT TGG-3`	The present study
	reverse	5`-GCG CCC TAC CTC GGT GGA ACC AG-3`	

*** Forward primer BIS3 is a 5´-UTR (Untranslated region) primer and is also specific for 5´-UTR of BIS1.**
**** Reverse primer BIS3 is a 3´-UTR primer and specific only for BIS3 isoenzyme.**

1.5.2.8 Primers used to clone SaPAL cDNA

Source	Primer	Sequence	Reference/ Accession number
Malus domestica	forward	5`-GAG TTT TCA GTT TTT CGT AAT TAA CAT-3`	CN903802
	reverse	5`-GTT GAA TGT GAA GGA ATG CAG C-3`	X68126
Pyrus communis	forward	5`-ATG GAG GCG GAG ACC ATC ACG C-3`	Fischer et al., 2007 / DQ230992
	reverse	5`-CTA ACA GAT AGG CAG AGG TGC G-3`	

1.5.3 Enzymes

Reverse transcriptases	RevertAid ^{TM} H Minus M-MuLV RT	Fermentas
	PrimeScript^{TM} Reverse transcriptase	TaKaRa Bio Inc.
DNA polymerases	*Taq*-DNA Polymerase (Dream *Taq*)	Fermentas
	Platinum® *Pfx* DNA Polymerase	Invitrogen
Endonucleases (restriction enzymes)	*Eco*R1, *Nhe*I, *Kpn*I, *Stu*I, *Xho*I	Fermentas
Other enzymes	T4-DNA Ligase, RNase A, RNase H, RNase inhibitor	Fermentas

1.5.4 Kits

RNA isolation	Rneasy® Plant mini kit	Qiagen
DNA isolation	Dneasy^{TM}plant maxi kit	Qiagen
Purification of DNA from PCR, restriction product, or from gel	innuPREP DOUBLE pure kit	Analytic Jena biosolutions
Sequencing	BigDye®Terminator v1.1 Cycle Sequencing Kit	Appl. Biosystems

1.6 Equipments

Equipment	Model	Company (Manufacturer)
Water purification system	Arium 611 VF	Sartorius, Germany
Autoclave	Vx-120	Systec GmbH Laborsystemtechnik
Balance	Small and large scale	Satorius,Germany
pH meter	Digital pH meter 325	WTW (wissenschaftlich-technische-werkstätten)
Centrifuge	Universal 32R	Hettich
	Biofuge 13	Heraeus Sepatech
	Sigma 1-15K	Sigma
GC-MS	Gas chromatograph 6890	Agilent
	ZB5-MS column (30 m, 0.25 mm i.d, 0.25 µm ft)	Phenomenex, Aschaffenburg
	Mass spectrometer	JEOL
LC-MS (ESI-MS)	3200 Q TRAP	Appl. Biosystems
Scintillation System	LS 6500	Beckman Coulter
Spectrophotometer	Ultrospect 1000	Pharmacia Biotech
Vacuum concentrator	RVC 2-18	CHRIST
Incubator shaker	HT	Infors
Thermo block	Dri-Block DB-3D	Techne
HPLC	Hewlett Packard HPLC system (1090 Series II) Dual absorbance detector	Agilent
HPLC Columns	**A**-Hypersil GOLD (5 µm; 150 mm, 4.6 mm)	Thermo Scientific
	B-Phenomenex C_{18} (3 µm; 150 mm, 3.2 mm)	Phenomenex
	C-Discovery HS F5 (3 µm; 150 mm, 2.1 mm)	Supelco
Balance	LA 230S	Satorius
Sequencer	ABI Prism	Appl.Biosystems
Gel documentation	MultiImage ^{TM} Light Cabient	Alpha In. Corp
Heating Circulator water bath	MW-4	Julabo
PCR cycler	T-Proffessional Gradient	Biometra
Clean bench	LaminarAir HLB 2472	Heraeus
	Laminar Air HBB 2460	Heraeus
Magnetic rotator	VF2	IKA-Labortechnik (Janke & Kunkel)
Electrophoresis	Mini-Sub Cell	BioRad
	Sub Cell-GT	BioRad
	Protein Chamber	Biometra
Power supply	Standard Power Pack P25	Biometra
	Power Pack 300	BioRad
Ultrasonic-Cell-Disruptor	Sonifier 250	Branson (G.Heinemann)
-80°C freezer	Hera Freeze	Heraeus

2 Biochemical methods

2.1 Stabilization of *in vitro* cultures

Callus of *Sorbus aucuparia* L. was derived from young shoots. After surface sterilization, stem segments were grown on solid LS medium to produce callus. The callus was transferred to 50 ml liquid LS medium and cultivated in the dark. The resulting cell suspension cultures [Figure 11] were shaken at 180 rpm and 24°C. Cultured cells (3 g) were transferred into 50 ml fresh medium every 14 days.

2.2 Elicitation of *in vitro* cultures

In a dose response study using 2.5-50 mg L^{-1} chitosan, a final concentration of 25 mg L^{-1} induced the maximum aucuparin level (Gaid et al., 2009). Five-day-old cell cultures from the linear growth phase were treated with chitosan at a final concentration of 25 mg L^{-1}. Nine and 18 h after the onset of treatment, cultured cells were used for nucleic acids (RNA or DNA) and crude protein extract preparation, respectively.

2.2.1 Chitosan preparation

Chitosan (a polymer of β-1,4-glucosamine residues) is an effective elicitor that is extensively used for stimulating plant secondary metabolite formation (Pitta-Alvarez and Giulietti, 1999; Komaraiah et al., 2002; Chakraborty et al., 2008).

The chitosan stock solution was prepared according to Pitta-Alvarez and Giulietti (1999). A concentrated chitosan solution was prepared by dissolving the polymer in 1% (v/v) acetic acid. The pH was adjusted to 5.5 with 1N NaOH and sterilized by autoclaving at 120°C and 1 atm for 20 minutes. The solution was kept at 4°C.

2.3 Enzyme extraction

Cell free extract used to measure benzaldehyde dehydrogenase activity was prepared according to Sircar and Mitra (2008). Soluble proteins were extracted from elicitor-treated cell cultures at 0-4°C. Cells (5 g) were collected by filtration, mixed with 10% (w/w) Polyclar AT (Serva, Heidelberg, Germany) and homogenized in 4 ml of 100 mM HEPES buffer pH 8.0 containing 10 mM dithiothreitol (DTT). The homogenate was centrifuged at 9000 rpm for 25 min at 4°C. An aliquot of the supernatant (2.5 ml) was freed from the low molecular mass substances by gel filtration through a PD_{10} column (GE Healthcare, Freiburg, Germany) equilibrated with 200 mM Tris-HCl buffer pH 7.5 (3.5 ml).

Cell-free extracts used to measure biphenyl synthase (BIS) activity were prepared as described by Liu et al. (2004), except that ammonium sulphate precipitation was omitted. Cultured cells (3 g) were mixed with 0.3 g polyclar AT and homogenized in 3 ml of 0.1 M potassium phosphate buffer (pH 7.5) containing 1 mM DTT. After centrifugation, the supernatant (ca. 2.5 ml) was passed through a PD_{10} column and equilibrated with the same extraction buffer. All steps were carried at 0-4°C. Polyclar AT helps to antagonize the action of oxidation by chelating polyphenolic substances.

Centrifugal devices provided rapid and convenient concentration, purification, and desalting of small volumes of biological samples. Diluted protein samples were concentrated by filtration using a highly selective ultra filtration membrane. This membrane was made from polyethersulphone specifically modified to minimize protein binding. The centrifugal device was selected to have a molecular weight cutoff (MWCO) 3 to 6 times less than the molecular mass of the protein to be purified. The driving force for filtration was provided by centrifugation at 3000 to 7500 g. The protein molecules larger than the MWCO of the membrane were retained in a sample reservoir, while solvent and low molecular mass proteins passed through the membrane into the filtrate reservoir. Vivaspin 6 Centrifugal Concentrators with patented vertical membrane technology were used for this purpose.

2.4 Determination of protein content

The total protein concentration was determined by binding of proteins to Coomassie® brilliant Blue G 250 (Bradford, 1976). In the photometric assay, the binding ability causes a shift in the absorbance from 465 nm to an intense band at 595 nm. In the assay, 10 µl sample was mixed with 90 µl buffer and 900 µl Bradford-dye solution (II.1.4.6). The assay was performed in a cuvette of one cm width at 595 nm in an UV/VIS spectrophotometer. Protein concentrations were calculated from a calibration curve, which was prepared for each determination using 1 to 10 µg /ml bovine serum albumin (BSA) as standard.

2.5 Benzaldehyde dehydrogenase (BD) assay

Benzaldehyde dehydrogenase activity was assayed essentially as described by Sircar and Mitra (2008). The standard assay mixture consisted of 0.5 mM benzaldehyde, 1 mM NAD^+, and 130 µg protein. The final assay volume was adjusted to 200 µl with 200 mM Tris-HCl buffer pH 9.5. The reaction mixture was incubated at 40°C for 30 min. Control assay contained boiled protein extract. The reaction was stopped by adding an equal volume of ice-cold methanol:acetic acid (9:1, v/v). After rigorous shaking, the sample was centrifuged at 10,000 g for 10 min and the resulting supernatant was analyzed by HPLC. Three independent experiments were performed and the average values were calculated.

2.6 Biphenyl synthase (BIS) assay

Biphenyl synthase activity was measured as described by Liu et al. (2004). The enzyme assay was performed in a total volume of 250 µl containing 15 µM benzoyl-CoA as starter CoA, 56 µM malonyl-CoA as extender CoA, 0.1 M potassium phosphate buffer pH 7.0 and approximately 100 µg protein. The mixture was incubated at 37°C for 30 min. The reaction was stopped by adding 20 µl 3 M trichloroacetic acid. The assay was extracted twice with 250 µl ethyl acetate and centrifuged at 13000 rpm for 10 min. The combined organic phases were dried under vacuum, and the residue was dissolved in 50 µl of methanol. Analysis of the enzymatic products was performed by HPLC. For each set of incubations, a positive control assay was done by using 2 µg recombinant BIS1 protein cloned in our group (Liu et al., 2007).

2.7 GC-MS analysis of BD assay

For GC-MS analysis, the BD assay volume was increased to 1 ml. The incubation was stopped by adding 100 µl of 3 M trichloroacetic acid and extracted with two volumes of dichloromethane. The organic phase in micro-vials was evaporated to dryness in a gentle stream of nitrogen gas. The residue was subjected to a two-step-derivatization procedure. For methoxymation of benzaldehyde, 10 µl of 20 mg/ml freshly prepared methoxyamine hydrochloride (Aldrich, Taufkirchen, Germany) in pyridine were added, followed by incubation for 30 min at 40°C under occasional shaking. Thereafter, TMS derivatization of benzoic acid was performed by addition of 40 µl of *N*-methyl-*N*-(tri-methylsilyl)-trifluoroacetamide (MSTFA; ABCR, Karlsruhe, Germany). After incubation for 30 min at 60°C, the solution was directly analyzed by GC-MS. GC-MS analysis was carried out using an Agilent 6890 gas chromatograph equipped with a ZB5MS column (30 m long, 0.25 mm i.d., 0.25 µm ft). Injector and transfer line were set at 250°C. The temperature program was 70°C (3 min) to 310°C (3 min) at 10°C min^{-1}. The split ratio was 1:10, the injection volume 1 µl and the carrier gas flow 1 ml min^{-1} He. The capillary column was directly coupled to a Jeol GC Accu TOF ^{Tm}. The retention index (RI) was calculated by a set of hydrocarbons (even numbered C_{12} to C_{28}) by linear interpolation.

2.8 HPLC analysis of BD assay

BD enzymatic products were prepared as described above. Supernatants were analyzed using a Hewlett Packard HPLC system (1090 Series II) coupled with a Dual Absorbance Detector and equipped with Chemstation (Pascal series) software on a Windows 95 platform.
Solvent gradients and columns were as follows:

<table>
<tr><th colspan="2">Gradients</th><th>Columns (see II.1.6)</th><th>To analyze assays containing the following substrates</th></tr>
<tr><td colspan="2">Isocratic mixture of water 55% and methanol 45% containing 1 mM TFA for 30 min.</td><td>Column A with flow rate 0.5 ml min^{-1}</td><td>Benzaldehyde
Cinnamaldehyde
2-Hydroxybenzaldehyde
2-Anisaldehyde</td></tr>
<tr><td colspan="2">Isocratic mixture of water 68% and methanol 32% containing 1 mM TFA for 30 min.</td><td>Column A with flow rate 0.5 ml min^{-1}</td><td>4-Hydroxybenzaldehyde
Vanillin
3,4-Dihydroxybenzaldehyde</td></tr>
<tr><td>Time (min)
0
5
15
28
30
35
40
45</td><td>Methanol (%)
5
5
40
60
100
100
5
5</td><td>Column B with flow rate 0.3 ml min^{-1}</td><td>3,4,5-Trihydroxybenzaldehyde</td></tr>
<tr><td colspan="3">Water and methanol contain 0.1% phosphoric acid</td></tr>
<tr><td>Time (min)
0
5
15
20
30
35
40
45</td><td>Acetonitrile (%)
5
5
40
60
100
100
5
5</td><td>Column C with flow rate 0.2 ml min^{-1}</td><td>3-Hydroxybenzaldehyde</td></tr>
<tr><td colspan="3">Water and acetonitrile contain 0.1% phosphoric acid</td></tr>
</table>

Detection wavelengths of BD enzymatic products together with their authentic references were as follows:

Substrate	Product	Detection wavelength of product λ (nm)
Benzaldehyde	Benzoic acid	227
2-Anisaldehyde	2-Anisic acid	
2-Hydroxybenzaldehyde	2-Hydroxybenzoic acid	
4-Hydroxybenzaldehyde	4-Hydroxybenzoic acid	254
Vanillin	Vanillic acid	
3,4,5-Trihydroxybenzaldehyde	Gallic acid	295
3-Hydroxybenzaldehyde	3-Hydroxybenzoic acid	
Cinnamaldehyde	Cinnamic acid	280
3,4-Dihydroxybenzaldehyde (Protocatechuic aldehyde)	3,4-Dihydroxybenzoic acid (Protocatechuic acid)	

The identity of the enzymatic products was confirmed by UV spectroscopy, co-chromatography with authentic reference compounds and by GC-MS for benzoic acid product in the standard assay.

2.9 HPLC analysis of BIS assay

Gradient		Columns (II.1.6)	Detection wavelength λ (nm)
Time (min)	Methanol (%)	Column B with flow rate 0.3 ml min^{-1}	281
0	50		
2	50		
12	70		
20	90		
21	95		
23	95		
25	50		
30	50		
Water and methanol containing 0.1% phosphoric acid			

The identity of the enzymatic products was confirmed by UV spectroscopy and co-chromatography (HPLC) with authentic reference compounds.

2.10 Time course changes in BD and BIS activities

Time-course changes in BD and BIS activities were measured in chitosan-treated cell cultures. Chitosan-treated (25 mg/L) cell suspension cultures (5-day-old) were harvested at different time points after the onset of elicitation (3-24 h). Cell-free extracts were prepared and used for the determination of BD and BIS activities at defined time points.

2.11 Biochemical characterization of BD

For biochemical characterization of BD enzyme, cell-free extract was prepared at 18 h after the onset of chitosan treatment. All incubations were performed in triplicate and average values were calculated.

A- Determination of pH and temperature optima

To study the pH optimum, 200 mM potassium phosphate buffer ranging from pH 5.5-6.5 and 200 mM Tris-HCl buffer ranging from pH 7.5-11.5 were used for optimum buffer capacity. At the optimum pH value, another series of incubations were performed at different temperatures between 15 - 50°C.

B-Linearity with protein amount and incubation time

The amount of enzymatically formed benzoic acid (nmol) was determined as a function of the protein amount in the standard assay (20 -320 µg) and incubation time (5, 10, 15, 30, 90, 120, 180, and 240 min).

C- Thiol reagent (DTT) dependence

BD activity was determined as a function of DTT concentration in the standard assay (0, 0.05, 0.01, 0.02, 1, and 10 mM).

D-Effect of Divalent metal ions on BD activity

Effect of divalent metal ions (Fe^{2+}, Mg^{2+}, Ca^{2+}, Mn^{2+}, Co^{2+}, Zn^{2+}, and Cu^{2+}) on BD activity was tested using the chloride salts at a final concentration of 1 mM.

E-Effect of cofactors on BD activity

NAD^+, $NADP^+$, FAD and FMN were tested as co-factor in BD reaction. All cofactors were used in a final concentration of 1 mM. Assay conditions were pH 9.5, 30 min incubation time, 40°C incubation temperature, and 130 µg protein per assay.

F-Study of substrate specificity

At the pH and temperature optima, BD assays were performed using a series of substrates (benzaldehyde, cinnamaldehyde, 2-hydroxybenzaldehyde, 3-hydroxybenzaldehyde, 4-hydroxybenzaldehyde, protocatechuic aldehyde, 2-anisaldehyde, vanillin, 3,4,5-trihydroxybenzaldehyde). All substrates were used in a final concentration of 0.5 mM with 1 mM NAD^+ in absence of DTT. Incubation time was 30 min with 130 µg protein per assay at 40°C.

G- Determination of kinetic parameters

The kinetic properties of BD were determined using different concentrations of benzaldehyde (0.01-1 mM) at a fixed concentration of NAD^+ (2 mM) or using different concentrations of NAD^+ (0.01-2 mM) at a fixed concentration of benzaldehyde (1 mM). The protein amount used was 50 µg per assay so that the reaction velocity was linear during the assay period. The apparent K_m values for the BD substrates were calculated from Hanes plots using Hyper 32, a hyperbolic regression programme for the analysis of enzyme kinetic data (http://homepage.ntlworld.com/john.easterby/software.html).

2.12 *p*-Coumarate-CoA ligase assay

Cinnamate and (hydroxy) cinnamate-CoA ligase activities were measured spectrophotometrically at room temperature with a 1 ml mixture containing 100 mM Tris–HCl, pH 7.5, 2.5 mM $MgCl_2$, 2.5 mM ATP, 0.6 mM (hydroxy) cinnamic acids and 0.2 mM CoA. Protein amount was 50 µl crude extract or 5 µg pure recombinant protein. The assay was started by the addition of CoA. The change in absorbance of the reaction mixture was monitored at the wavelengths of 311, 333, 345, 346, and 352 nm according to the reported absorption maxima for cinnamoyl-CoA, *p*-coumaroyl-CoA, feruloyl-CoA, caffeoyl-CoA, and sinapoyl-CoA, respectively (Gross and Zenk, 1966; Stöckigt and Zenk, 1975).

Enzymatic product	**ε ($mM^{-1}cm^{-1}$)**
p-Coumaroyl-CoA	21
Caffeoyl-CoA	18
Feruloyl-CoA	19
Cinnamoyl-CoA	22
Sinapoyl-CoA	20
mM product = Absorbance at specific λ / Extinction coefficient For accurate results, absorbance should be in the range from 0.15-0.85	

2.13 Benzoate-CoA ligase assay

Radioisotopic assays were performed with 100 µl buffer containing 100 mM Tris–HCl, pH 7.5, 2.5 mM $MgCl_2$, 2.5 mM ATP, 0.2 mM CoA, 50000 cpm [7-^{14}C] benzoic acid and 5 µg pure protein. The assay was started by the addition of CoA and kept at room temperature for 30 min. The reaction was stopped by the addition of 20 µl 3 M trichloroacetic acid and extracted twice with 100 µl ethyl acetate, vortexed, and phase separated by a 1-min centrifugation at 13,000 rpm. The upper organic phase was removed and the ethyl acetate extraction was repeated. The level of radioactivity in the remaining aqueous phase was counted in a liquid scintillation counter, Beckman LS 6500. The amount of radioactivity in the aqueous phase indicated the amount of synthesized benzoyl-CoA.

2.14 Purification of enzymatic *p*-coumaroyl-CoA

CoA esters were purified using solid-phase extraction (SPE) LC-18 column (Supelco, 1 g) preconditioned with consecutive washes of methanol, dH_2O, and 4% ammonium acetate solution (5x column volume each). One ml incubation (II.2.12) was mixed with equal volume of 8% ammonium acetate to a final concentration of 4% and the mixture was loaded onto the column. The column was rinsed with 4% ammonium acetate solution until the flow-through showed the absence of free CoA (determined by spectrophotometry). The CoA esters were recovered by elution with distilled water (Beuerle and Pichersky, 2002). Fractions containing *p*-coumaroyl-CoA were analyzed *via* ESI-MS (II.2.15).

2.15 Electrospray ionization-mass spectrometry (ESI-MS)

Purified *p*-coumaroyl-CoA (0.1 mM) (II.2.14) was analyzed and characterized *via* ESI MS/MS. The purified solution was analyzed by direct infusion with a flow rate of 5-10 µl/min. Mass spectrum analysis was carried out through "Positive Full-Scan Mode". MS/MS-experiment was done with EPI+ (enhanced product ion scan, positive mode). Instrument tuning and mass calibration was based on molecular ion peak $[M+H]^+$ through the whole fragmentation pattern.

Instrumental

Mass spectrometer	Applied Biosystems 3200 Q TRAP, Turbo V Ion Source
Software	Analyst 1.4.2
Collision gas	N_2

Parameter/Experiment	**Q1+**	**EPI+**
Curtain gas	10 ml/min	10 ml/min
Collision gas	-	medium
Ion spray voltage	5500 V	5500 V
Heater	-	-
Ion source gas 1	10 ml/min	10 ml/min
Ion source gas 2	-	-
Declustering potential	61V	73 V
Entrance potential	11 V	12 V
Collision energy	-	39 V
Collision cell exit potential	-	4 V

2.16 Biochemical characterization of 4CL

For biochemical characterization of 4CL enzyme, pure recombinant protein was used. All incubations were performed in triplicate and average values were calculated.

2.16.1 Determination of pH and temperature optima

To study the pH optimum, 100 mM potassium phosphate buffer ranging from pH 5.5-6.5 and 100 mM Tris-HCl buffer ranging from pH 7.5-11.5 were used for optimum buffer capacity. At the optimum pH value, another series of incubations were performed at different temperatures between 25–55°C

2.16.2 Linearity with protein amount and incubation time

The amount of enzymatically formed *p*-coumaroyl-CoA (nmol) was determined as a function of the protein amount in the standard assay (1 -200 µg) and the incubation time (from 2 till 60 min).

2.16.3 Effect of divalent metal ion on 4CL activity

Effect of divalent metal ions (Mg^{2+}, Co^{2+}, Mn^{2+}, Fe^{2+}, Ca^{2+}, Zn^{2+}, and Cu^{2+}) on 4CL activity was tested using the chloride salts at a final concentration of 2.5 mM.

2.16.4 Study of substrate specificity

At the pH and temperature optima, 4CL assays were performed using a series of substrates (*p*-coumaric, caffeic, ferulic, cinnamic, and sinapic acids). All substrates were used at a final concentration of 0.6 mM. Protein amount was 5 µg per assay.

2.16.5 Determination of kinetic parameters

At optimum pH, temperature, and incubation time, the kinetic properties of 4CL were determined using different concentrations of substrates (1-600 µM). The ATP and CoA ranges used were 1-200 µM. Appropriate enzyme concentration (2 µg) was chosen so that the reaction velocity was linear during the assay period. The K_m values for the 4CL substrates were calculated from Lineweaver-Burk plots using Hyper 32, a hyperbolic regression programme for the analysis of enzyme kinetic data (http://homepage.ntlworld.com/john.easterby/software.html).

3 Molecular biology methods

3.1 Design of gene specific primers

In general, a primer should contain a minimum of 18-25 nucleotides complementary to the sequence of interest and have a GC content of approximately 45-55%. If the introduction of a restriction site is required, it should be flanked by 3-10 (depending on the enzyme) "spacer" nucleotides at the 5´ end of the primer to allow an efficient digestion close to the end of the cDNA as in case of overexpression primers (II.1.5.2.3).

3.2 Isolation of total RNA and genomic DNA

Isolation and purification of total RNA and genomic DNA were performed using the RNeasy® Mini Kit and DNeasy Plant Maxi Kit (Qiagen), respectively. As per manufacturer's instruction, 100 mg freshly harvested cell suspension culture was ground with liquid nitrogen and used as a material for extraction. Using silica-gel membrane technology, U.S. Patent No. 5,234,809 describes a procedure to isolate nucleic acids from biological samples, which uses a chaotropic agent together with a silica based nucleic acid binding solid phase. Guanidine hydrochloride at pH 3 to 5 or guanidine thiocyanate at higher pH, combined with other salts, is used as the chaotropic agent. After binding of the DNA to the solid surface, the solid phase washed with the chaotropic agent to remove any biological contamination followed by treatment with 70% ethanol to remove the chaotrope. The DNA is eluted using water or 10 mM Tris-HCl pH 8.

Purified RNA was used either directly for reverse transcription or stored at -80°C. Genomic DNA can be stored with relatively high stability at -20°C.

3.3 Quantification of RNA and DNA concentration

The concentration of nucleic acids was determined by measuring the absorbance value of the DNA or RNA samples at wavelength of 260 nm. One unit absorbance value at wavelength 260 nm (ε_{260}) corresponds to 40 ng/µl single-stranded RNA or 50 ng/µl double-stranded DNA (Sambrook et al., 2001).

$[c] = A^* \varepsilon_{260}^*$ dilution factor [c]; concentration of RNA or DNA.

Purity is determined by calculating the ratio of absorbance at 260 nm to absorbance at 280 nm. Pure samples having an A_{260}/A_{280} ratio of 1.7-1.9 ensure the appropriate quality.

Contaminations with proteins or phenolics would reduce this value due to their absorbance at 280 nm.

3.4 Reverse transcription (RT)

Reverse transcriptase is an RNA-dependent DNA polymerase and used to catalyze the oligo-dT primed synthesis of first strand cDNA (DNA complementary to the appropriate mRNA) from total RNA**.** After reverse transcription, RNA template is degraded by incubation with RNaseH at 37°C for 20 min. RNaseH is an endoribonuclease that specifically hydrolyzes the 5'-phosphodiester bonds of RNA, which is hybridized to DNA. The reverse transcription reaction product can be directly used or stored at -20°C.

3.4.1 cDNA-synthesis using SMART-RACE protocol

PrimeScript™ Reverse transcriptase is based on reverse transcriptase from M-MuLV (Moloney Murine Leukemia Virus). It exhibits a terminal transferase activity by adding 3-5 cytosine (C) residues to the 3´ end of the first strand cDNA. SMART II- primer contains a terminal stretch of guanine (G) residues that anneal to the dC-rich cDNA tail and serves as an extended template for RT.

Following reverse transcription, the first-strand cDNA is used directly in 5´-RACE PCR reactions together with reverse gene-specific primers (II.1.5.2.1).

5´CDS-cDNA synthesis reaction

Component	**Volume**
1µg total RNA	x µl
Smart II oligo primer (10 pmol/µl)	1 µl
5'-CDS primer (10 pmol/µl)	1 µl
DTT (0.1M)	2 µl
Nuclease free water up to	10 µl
Denaturation of RNA at 70°C for 10 min followed by briefly cooling on ice and addition of the following reagents	
5x PrimeScript™ buffer	4 µl
10 mM dNTP mix	1 µl
The mixture was prewarmed at 37°C for 5min followed by addition of	
PrimeScript™ Reverse transcriptase (200 U/µl)	0.5 µl
Nuclease free water up to	20 µl
The reaction mixture was incubated at 42°C for 90 min, followed by heat inactivation of the enzyme at 70°C for 15 min.	

3.4.2 cDNA-synthesis using H Minus M-MuLV RT

3'-RACE cDNA is synthesized using the reverse transcriptase enzyme "H Minus M-MuL V RT" (Fermentas), which lacks terminal transferase activity. The 3'-RACE cDNA is synthesized using a traditional reverse transcription procedure, but with a special 3'-RACE CDS primer (3'CDS). This 3'CDS primer includes a portion of SMART-II primer sequence at its 5'end. Following reverse transcription, the first-strand cDNA is used directly in 3'-RACE PCR reactions together with forward gene-specific primers (II.1.5.2.2).

3'CDS-cDNA synthesis reaction

Component	**Volume**
1µg total RNA	x µl
Oligo dT primer (3'CDS) 10 pmol	1 µl
RNase inhibitor (40 U/µl)	0.5 µl
Nuclease free water up to	10 µl
Denaturation of RNA at 70°C for 10 min followed by briefly cooling on ice and addition of the following reagents	
5x reaction buffer	4 µl
10 mM dNTP mix	1 µl
The mixture was prewarmed at 37°C for 5 min followed by addition of	
H Minus M-MuL V RT(200 U/µl)	1 µl
Nuclease free water up to	20 µl
The reaction mixture was incubated at 42°C for 90 min, followed by heat inactivation of the enzyme at 70°C for 15 min.	

3.5 Polymerase Chain Reaction (PCR)

PCR is carried out by using two oligonucleotide primers that flank the DNA fragment to be amplified. These primers hybridize the opposite strands of the target sequence and are oriented so that DNA synthesis proceeds by the polymerase across the region between the primers. This segment is doubled in every cycle of the PCR resulting in an exponential accumulation of the specific fragment.

Nucleotide sequence information obtained from 4CL1 (Scharnhop, 2008) enabled the design of reverse Sa4CL1 gene-specific primers, Rev-P1-P4 (II.1.5.2.1). Forward gene-specific primers (II.1.5.2.2) were also designed from the sequence data obtained for 4CL4. Designed GSP in combination with adapter primers, RACE Long, RACE Short (Clontech), were used in attempt to identify the 5′ end of Sa4CL1-cDNA and 3′-end of 4CL4.

Standard PCR

Component	Volume
DNA (1 µg)	1 µl
Forward primer (10 pmol)	1 µl
Reverse primer (10 pmol)	1 µl
10x reaction buffer with 20 mM $MgCl_2$	2.5 µl
dNTPs (10 mM)	1 µl
Polymerase (5 U/µl)	0.1 µl
Nuclease free water up to	25 µl

Standard PCR programme

Step	Temp [°C]	Time [sec]	Cycles	Aim
1	94	180		Denaturation
2	70	Pause		Hot start
3	94	45		Denaturation
4	T_m *	45		Annealing
5	72	120	30x from step 3	Elongation
6	72	600		Final elongation
7	10	Pause		stop

In case of *pfx* proofreading DNA polymerase:
Denaturation time in step 3 decreased from 45 sec to 30 sec.
Elongation temperature was decreased to 68°C in step 5 and 6
Elongation time increased to be 240 sec in initial elongation and 1200 sec in final elongation.
T_m = 55.5°C

*T_m = 2°C (A+T) + 4°C (G+C)

3.5.1 Touch down PCR

Touch down PCR uses an annealing temperature during the initial PCR cycles that is 5-10°C higher than melting temperature (T_m) of the universal primer. In subsequent cycles, the annealing temperature is decreased in increments of 0.5°C per cycle until the temperature was reached 5°C below the T_m of the GSP.

Touch down PCR programme

Step	Temp [°C]	Time [sec]	Cycles
1	94	180	
2	70	Pause	
3	94	45	
4	T_m+5°C	45	T_m decrease 0.5 degree each cycle
5	72	60	10x from step 3
6	94	45	
7	T_m	45	
8	72	120	30x from step 6
9	72	600	
10	15	pause	

3.5.2 Rapid amplification of cDNA ends (RACE)

RACE is a polymerase chain reaction-based technique developed to obtain the full length cDNA by extending a cDNA fragment toward 3´- or 5´-ends. RACE technique allows the identification of cDNA from a messenger RNA template between a defined internal site and unknown sequence at either 3´ or 5´ ends of the mRNA.

3´-RACE takes advantage of the natural poly (A) tail in mRNA as a generic priming site for PCR amplification. In this procedure, mRNAs were converted into cDNAs using reverse transcriptase (RT) and an oligo-dT adapter primer (3'CDS primer).

The first strand cDNA synthesized as described in II.3.4.2 served as template to amplify the target sequence using a GSP (II.1.5.2.2) that anneals to a region of known sequence and an oligo-dT universal primer that targets the poly A tail region.

To permit the amplification of cDNA by 5´ RACE in a similar manner as 3' RACE, an artificial poly dC tail had to be added first to the 3´ end of the 1st strand cDNA (through a reverse transcriptase with terminal transferase activity). The resulting cDNA (II.3.4.1) could be amplified by using a reverse GSP (II.1.5.2.1) and poly dG-containing anchor primer (SMART II primer, II.1.5.2.5).

3.5.3 TATA-box protocol

(Guo et al., 2010)

The core promoter region can extend ~35 bp upstream the transcription start site. A minimal promoter region required to start the pre-initiation complex formation usually has a TATA-box, which is conserved in most of the species (30–50% of promoters), and a transcription start site (TSS) region, which is usually not conserved.

The TATA-box is conserved in the promoter region of genes in eukaryotic organism. Guo et al. (2010) designed the TATA-box degenerate (TD) primer (II.1.5.2.6) according to TATA-box nucleotide frequency matrices (Shahmuradov et al., 2003). At the 5´-terminus of TD primers, nine nucleotides have been added as a 5´ clamp to decrease the degeneracy of TD primers and to improve the PCR round. One µl (per 25 µl reaction) of first round PCR product was used as the template of second round PCR (25 dilutions), while 25 dilutions of second round PCR product served as template of third round PCR. Reverse gene-specific primers, complementary to a known coding region and with higher annealing temperature (TA) than TD primers, have been designed (II.1.5.2.1)

TATA-box PCR

Component	Volume
DNA (1 µg)	5 µl
TD primer (50 pmol)	1 µl
Reverse primer GSP (10 pmol)	1 µl
10x reaction buffer with 20mM $MgCl_2$	2.5 µl
dNTPs (10 mM)	1 µl
Polymerase (5 U/µl)	0.25 µl
Nuclease free water up to	25 µl

Cycling conditions used for semi-nested PCR with TD primers

Reaction	Step	Temp. (°C)	Time (Sec)	Cycles
First amplification	1	98	180	
	2	70	Pause	
	3	95	30	
	4	62	30	
	5	72	150	5x from step 3
	6	95	30	
	7	30	40	
	8	72	150	
	9	95	30	
	10	58	30	
	11	72	150	4x from step 9
	12	95	30	
	13	56	30	
	14	72	150	9x from step12
	15	72	300	
	16	12	pause	
Second and third amplifications	1	98	180	
	2	70	Pause	
	3	95	30	
	4	65	30	
	5	72	150	3x from step 3
	6	95	30	
	7	62	30	
	8	72	150	3x from step 6
	9	95	30	
	10	59	30	
	11	72	150	3x from step 9
	12	95	30	
	13	56	30	
	14	72	150	15x from step 12
	15	72	300	
	16	12	pause	

3.6 Agarose gel electrophoresis

Agarose gels were prepared at concentrations of 0.8-2% agarose in electrophoresis buffer (1x TAE, II.1.4.1). The agarose concentration was determined according to the expected band size. Increasing the agarose concentration of a gel reduces the migration speed and enables separation of smaller DNA molecules. The mixture was boiled in a microwave oven to dissolve the agarose. After cooling down the mixture to 60°C, ethidium bromide (0.5 µg/ml) was added and the solution was poured into the gel tray. Ethidium is a flat molecule, just the right size to get in between the stacked base-pairs of double-stranded DNA. This, and its fluorescence, is what make it a sensitive dye for detecting DNA. Since ethidium bromide is a carcinogen, all steps were performed under the fume hood.

3.7 Purification of DNA from agarose gels

After gel electrophoresis (II.3.6), DNA fragments had to be eluted from gel for further applications such as restriction or ligation. Purification of DNA fragments was performed by using the Innu PREP DOUBLE Pure Kit (Analytic Jena biosolution) to obtain clean DNA fragment. DNA was adsorbed to silica gel membrane in the presence of high concentration of chaotropic salts. Contaminants were washed out and the DNA was eluted with a low ionic strength elution buffer or water.

3.8 Cloning of PCR product

3.8.1 Cloning into pGEM-T Easy vector

pGEM-T Easy vector allows the direct cloning of *Taq*-polymerase amplified PCR products. It is a 3 kb linearized plasmid, prepared by cutting the vector with *EcoR*V and adding a 3’ terminal thymidine to both ends. *Taq*-polymerase is a non-proofreading DNA polymerase, which adds a single 'A' overhang to each 3'-end of the PCR product. The resulting PCR product can be directly ligated into a linear vector with a 3´ terminal 'T' at both ends. The vector contains T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme *β*-galactosidase. Insertional inactivation of the α-peptide allows recombinant clones to be directly identified by color screening on indicator plates. Insert can be released by single digestion using one restriction enzyme (EcoRI, BstZI or Not I).

3.8.2 Cloning of *p*-coumarate-CoA ligase into expression vectors

The ORF of 4CL1 cDNA was re-amplified by PCR using Platinum *Pfx* Polymerase (Invitrogen) and 5'CDS-cDNA as a template. Sa4CL1 overexpression primers (II.1.5.2.2) were used to introduce *Nhe*I and *Kpn*I restriction sites (in case of pRSET B expression vector) or *Kpn* I only (in case of pGEX-G expression vector). Restriction sites are part of the vector MCS and should be absent from target insert sequence.

4CL1-overexpression primers	**Expression vectors (II.1.5.1.2)**
Sa.ExpX1-5` ***Nhe*I** & Sa.Exp.X1-3`***Kpn*I**	For pRSET B expression vector
Exp.4CL-X1-ATG & Sa.Exp.X1-3`***Kpn*I**	For pGEX-G expression vector

The resulting fragment was digested with *Nhe*I and *Kpn*I and then ligated into the *Nhe*I/*Kpn*I linearized expression vector pRSET B. In case of pGEX-G expression vector, the restriction sites used were *Stu*I (forward) and *Kpn*I (reverse) resulting in 3`- blunt end and 5`- sticky end. The resulting fragment was digested with *Kpn*I only and then ligated into *Stu*I /*Kpn*I linearized expression vector pGEX-G.

pRSET B and pGEX-G vectors were designed for expression of His-tag and GST-fusion protein in *E.coli*, respectively. DNA insert was positioned downstream and in frame with a sequence that encodes an N-terminal fusion peptide. This sequence includes an ATG translation initiation codon and permits the use of affinity chromatography for the purification of fusion protein.

3.8.3 Standard restriction reaction

The following reaction mixture was used as a standard restriction protocol

Component	**Volume**
0.2 µg purified DNA insert or vector	x µl
Endonucleases	0.5 µl
10x reaction buffer	2 µl
Nuclease free water up to	20 µl
Reaction was incubated at 37°C for 2 h	

Following digestion, it was usually worthwhile to gel-purify the vector prior to the ligation to remove uncut, residual nicked and supercoiled plasmid, which was transformed very efficiently relative to the desired ligation product.

3.8.4 Standard ligation reaction

The following reaction mixture was used as a standard ligation protocol

Component	Volume
Digested vector 50 ng	1 µl
Digested insert 50 ng	6 µl
10x ligation buffer	1 µl
T4-DNA-Ligase (5U/ µl)	0.5 µl
Nuclease free water up to	10 µl

In case of pGEM-T vector, reaction was incubated at 4°C overnight.
In case of pRSET B vector, reaction was incubated at 16°C overnight.
In case of pGEX vector, insert to vector ratio is 5:1 and amount of ligase enzyme was doubled. Incubation temperature was 16°C for 10 h, then lowered to 4°C for another 10 h.

3.8.5 Restriction analysis

This method was used for screening of recombinant plasmids that lack the *lac Zα* gene for blue/white selection (II.3.8.2). Plasmid DNA obtained from minipreparations (II.3.11) was digested with an appropriate restriction enzyme at the optimal condition suggested by the manufacturer. The choice of that enzyme depends on the knowledge of restriction sites within the DNA sequence of interest and the vector used for cloning.

3.9 Preparation of competent cells

E. coli cells treated with calcium ions can take in the foreign DNA-plasmid and are called "competent cells". Competent cells are used in the recombinant DNA experiments for transformation. Preparation was done according to modified methods (Mandel and Higa 1970; Cohen et al., 1972; Dagert and Erlich, 1979). A single colony from an overnight agar plate or a sterile loopful from frozen culture was inoculated into 5 ml of LB medium and incubated with shaking at 37°C overnight. One ml of this fresh culture was transferred to 50 ml LB medium and incubated at 37°C with shaking until the OD_{600} reached 0.6-0.8. The culture was then chilled for 10 min on ice, and the cells were harvested by centrifugation at 3000 rpm and 4°C for 10 min. The pellet was washed with 50 ml of ice-cold 0.05 M $CaCl_2$. Cells were harvested by centrifugation as above and resuspended in 20 ml of 0.05 M $CaCl_2$. The cell suspension was then kept in ice for 20 minutes. After centrifugation, pellets were resuspended in 940 µl of 0.05 M $CaCl_2$ solution. After 15 min incubation on ice, 140 µl glycerol was added. Afterwards, the competent cells were stored at -80°C as 50 µl aliquots until further use.

3.10 Transformation of plasmid DNA into *E.coli*

3.10.1 Transformation into DH 5α

A mixture of 5 µl foreign plasmid (II.3.8.1 and II.3.8.2) and 50 µl competent cells (II.3.9) was incubated at 4°C for 30 min. DNA can be forced into the cells by 45 seconds heat shock at 42°C followed by 5 minutes incubation at 4°C. In order to maximize transformation efficiency, cells were allowed to grow in 250 µl SOC medium with continuous shaking at 37°C for 1h. The cells are finally plated on LB-agar medium containing ampicillin (100 µg/ml). In case of pGEM-T Easy vector, blue/ white colony selection is used to differentiate between white colonies, which contain a DNA fragment in the cloning site of the plasmid, from the blue colonies, which contain the original plasmid without interruption of *lacZα* gene by foreign DNA. This is based on the ability of the *β*-galactosidase enzyme to cleave the artificial chromogenic substrate 5-bromo-4-chloro-3-indolyl-*β*-D-galactopyranoside (X-gal) added to LB plates (II.1.3.2). *β*-Galactosidase is expressed by α complementation of the amino-terminal fragment of *β*-galactosidase encoded by the vector, and carrying the *lacZα* gene, and the carboxy-terminal portion of *β*-galactosidase encoded by the genome of host cell. Introduction of a cloned DNA fragment into the multiple cloning site (MCS) of the vector, which is embedded in the coding region of *lacZα* gene, disrupts the amino-terminal fragment of *β*-galactosidase and prevents formation of active *β*-galactosidase enzyme resulting in white colonies.

3.10.2 Transformation into BL 21 (DE3) pLysS

Plasmid-containing DNA for heterologous protein expression (II.3.8.2) was transformed into the expression host *E. coli* BL21 cells. BL 21 is the strain of choice for high-level gene expression and production of recombinant protein in a bacterial system. The transformation procedure was the same as with DH 5α (II.3.10.1), except that the heat shock time was for 20 seconds and 700 µl SOC medium was used for growing bacterial culture. After one hour of incubation at 37°C, cells were harvested by centrifugation at 5000 rpm for 5 min. Supernatant (500 µl) was discarded and the pellets were resuspended in the rest of the medium. About 80 µl of bacterial suspension was finally plated on LB-agar medium containing ampicillin (100 µg/ml) and chloramphenicol (60 µg/ml). Ampicillin resistance is acquired from the expression vector and chloramphenicol resistance is acquired from the host cells. Bacterial cultures containing plasmids with target insert were stored at -80°C in autoclaved solution of 20%

glycerol /LB-medium. This continuous culture from *E. coli BL21(DE3)* containing the desired gene can be used as a permanent supply to start protein expression.

3.11 Isolation of plasmid DNA by minipreparation

Circular plasmid can be isolated from *E. coli*-DH5α by the alkaline lysis method developed by Bimboin and Doly (1979). An overnight culture was prepared by inoculating one colony from a growing plate into 5 ml liquid LB medium containing the appropriate antibiotic. An overnight culture (2 ml) was harvested by centrifugation at 5000 rpm for 5 min. The bacterial pellet was then resuspended into ice-cold 300 µl buffer 1 containing RNase A (II.1.4.5). For alkaline lysis with SDS, buffer 2 (300 µl) was added to the bacterial suspension followed by gentle mixing and incubation at room temperature for 5 minutes. To precipitate protein and large fragments of genomic DNA, ice-cold 300 µl buffer 3 was added, followed by gentle mixing and incubation on ice for 20 minutes, then centrifugation at 13,000 rpm for 10 minutes. The clear supernatant was extracted with chloroform. The plasmid DNA was precipitated from the aqueous phase by centrifugation with 2-propanol for 30 min at 13,000 rpm. The pellets were washed with 70% ethanol. The plasmid DNA was finally dried at 37°C and redissolved in distilled water (50 µl). This purified plasmid DNA can be used for transformation (II.3.10) or for restriction analysis (II.3.8.5).

3.12 Heterologous expression of recombinant protein in *E. coli*

An *E. coli* BL21-Codon Plus (DE3) colony containing the desired heterologous gene was used for production of recombinant protein encoded by cDNA that was cloned into expresssion vector (II.1.5.1.2). One colony (II.3.10.2) was inoculated into 10 ml of LB medium containing ampicillin (100 µg/ml) and chloramphenicol (60 µg/ml). Inoculated LB medium was incubated at 200 rpm and 37°C overnight. About 4 ml of the overnight culture was transferred to 100 ml LB medium and incubated at the same growing conditions. At an OD_{600} of 0.6-0.8, IPTG (1 mM) was added and the incubation temperature was reduced to 25°C. After 4 h of incubation, the cells were harvested by centrifugation and could be kept at -20°C for the next step (II.3.13).

3.13 Extraction and purification of expressed proteins from *E. coli* cells

Heterologously expressed protein was extracted from *E. coli* cells to determine its biochemical activity. Mechanical disruption of the cell membrane (sonication) is a common method for breaking the cell wall and isolating the soluble protein. The frozen cell pellet from 100 ml culture (II.3.12) was re-suspended in ice-cold 3 ml of 0.1 M potassium phosphate buffer, pH 7-7.5 (in case of GST-fusion protein; II.1.4.2.1) or in 3 ml lysis buffer with 20 mM imidazole pH 8 (in case of His_6-tagged protein; II.1.4.2.2). Sonication of the cells was carried out on ice for 5 min at 50% pulses using a Branson Sonifier B15 (Heinemann, Schwäbisch Gmünd, Germany). After centrifugation at 10,000 g and 4ºC for 10 min, the supernatant was stored on ice until it was applied to the affinity purification system (II.1.4.2.1). An aliquot of the supernatant had been taken as expressed protein control and stored at -20°C until SDS-PAGE (II.3.14).

In case of GST-fusion protein, glutathione slurry (200 µl) was added to the clear lysate (3 ml). After mixing gently by shaking (200 rpm on a rotary shaker) at 4°C for 1 h, the lysate-glutathione mixture was poured into an open column. The mixture was washed four times with 1 ml PBS buffer each (II.1.4.2.1). The GST-fusion protein was eluted using 3.5 ml elution buffer. An aliquot of the affinity purified protein (2.5 ml) was freed from glutathione (used for elution) by gel filtration through a PD_{10} column equilibrated with 0.1 M Tris-HCl, pH 7.5. Pure protein was mixed with 20% autoclaved glycerol and could be stored at -20°C.

In case of His_6-tagged protein, heterologously expressed protein was extracted from *E. coli* by the same way as GST-tagged protein. The purification procedure was performed on a nickel-nitrilotriacetic acid "Ni-NTA" protein purification system. Ni-NTA slurry (200 µl) was added to 3 ml of the cleared lysate. After shaking at 4°C for 1 h, the mixture was loaded into a column. Affinity purification started with 4 ml washing buffer (four times 1 ml each) (II.1.4.2.2). The His_6-tagged-fusion protein was eluted using 3.5 ml elution buffer. Imidazole (used for elution) was removed from the eluate by gel filtration through a PD_{10} column equilibrated with 0.1 M Tris-HCl pH 7.5 buffer.

3.14 SDS-PAGE

SDS-PAGE was used to confirm the successful expression of cloned cDNAs in heterologous expression systems (II.3.12). Proteins are loaded with SDS and become equally negatively charged. In addition, a thiol reagent disrupts their subunits. While migrating towards the anode, they are separated according to their molecular mass (Laemmli, 1970). The concentration of acrylamide and bisacrylamide in the separating gel was 12%, which allowed the highest resolution of proteins between 10 and 200 kDa. Protein samples to be analyzed were mixed with protein loading buffer (II.1.4.1) in a 1:1 ratio and denatured at 95°C for 5 min. To estimate the molecular mass of the separated protein, a prestained 10-170 kDa protein marker was loaded in parallel with the samples. The running conditions were 25 mA in the stacking gel, 35 mA in the separating gel and 200 V supplied by a Standard Power Pack P25 (Biometra). To check for successful protein expression, the gel was incubated overnight at room temperature in Coomassie blue staining solution (II.1.4.1), followed by destaining solution until clear bands appeared.

3.15 DNA sequencing

For sequencing plasmid DNA, after purification with alkaline lysis method (II.3.11), the ABI PRISM 377 (Applied Biosystems) DNA sequencer (Genetic Analyzer) was used. Nucleotide sequences were determined from both strands using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). The primers used to read the sequenced plasmid are listed in II.1.5.2.4

The chain-termination method developed by Sanger and coworkers has become the method of choice for sequencing DNA samples. The key principle of the Sanger method (Sanger et al., 1977) was the use of dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators. The classical chain-termination method requires a single-stranded DNA template, a DNA primer, a DNA polymerase, radioactively or fluorescently labelled nucleotides, and modified nucleotides that terminate DNA strand elongation. The chain-terminating nucleotides lack the 3'-OH group, required for the formation of a phosphodiester bond between two nucleotides, thus terminating DNA strand extension and resulting in DNA fragments of varying length.

Dye-terminator sequencing utilizes labeling of the chain terminator ddNTPs, which permits sequencing in a single reaction. In dye-terminator sequencing, each of the four dideoxynucleotide chain terminators is labeled with fluorescent dyes, each of which with different wavelengths of fluorescence and emission.

3.16 Computer-assisted sequence analysis

DNA sequences were analyzed using the EditSeq, MegAlign, SeqMan, and MapDraw options from the DNASTAR software. BLAST (Basic local alignment tool) is a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. This program is accessible at: www.ncbi.nlm.nih.gov/BLAST/ .

III. Results

1 Detection of benzaldehyde dehydrogenase (BD) and biphenyl synthase (BIS) activities

S. aucuparia cell cultures accumulate aucuparin when treated with chitosan, which is a well-known elicitor (Villegas and Brodelius 1990; Chakraborty et al., 2008). In a dose-response study using 2.5-50 mg/L chitosan, a final concentration of 25 mg/L induced the maximum aucuparin level (Gaid et al., 2009). Chitosan-treated cell cultures were used to prepare cell-free extracts for the incubation with benzaldehyde and NAD^+. BD from *S. aucuparia* cell suspensions catalyzed the *in vitro* conversion of benzaldehyde into benzoic acid. Subsequent HPLC analysis demonstrated the formation of benzoic acid [Figure 12]. The identity of this enzymatic product was confirmed by co-chromatography with a sample of authentic reference compound and UV-spectroscopy. In addition, GC-MS analysis revealed that the mass spectrum of the enzymatic product agreed with that of the reference compound [Figure 13]. No enzymatic benzoic acid formation was observed in control assays containing heat-denatured cell-free extract.

Incubation of cell-free extracts from chitosan-treated *S. aucuparia* cell cultures with benzoyl-CoA and malonyl-CoA led to the formation of the BIS product, as shown by HPLC analysis of the ethyl acetate extracts from the enzyme assays [Figure 14]. The enzymatic product was identified as 3,5-dihydroxybiphenyl by comparison with a sample of chemically synthesized reference compound (Liu et al., 2004). A similar incubation was done using affinity-purified recombinant BIS1 protein derived from a cDNA cloned in our group (Liu et al., 2007).

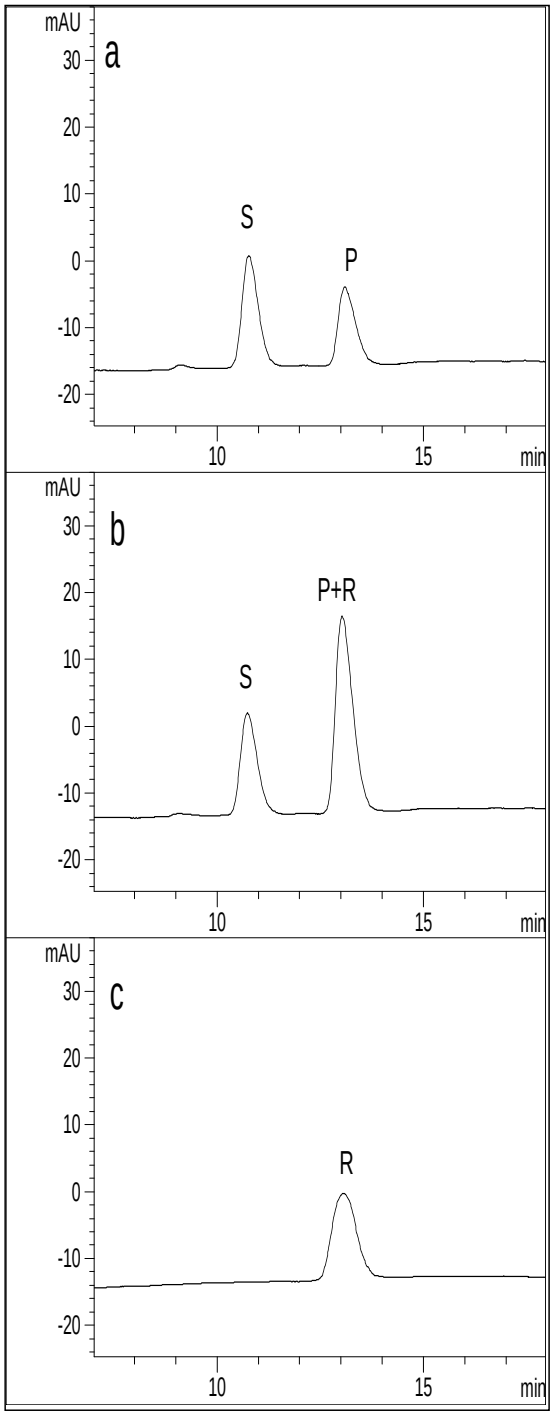


Figure 12. HPLC analysis of benzaldehyde dehydrogenase assays using cell-free extracts from cell cultures of *S. aucuparia,* (a) standard incubation, (b) co-chromatography of standard incubation and reference compound, (c) benzoic acid as reference. S, substrate (benzaldehyde); P, enzymatic product (benzoic acid); R, reference (benzoic acid). Detection wavelength, 227 nm.

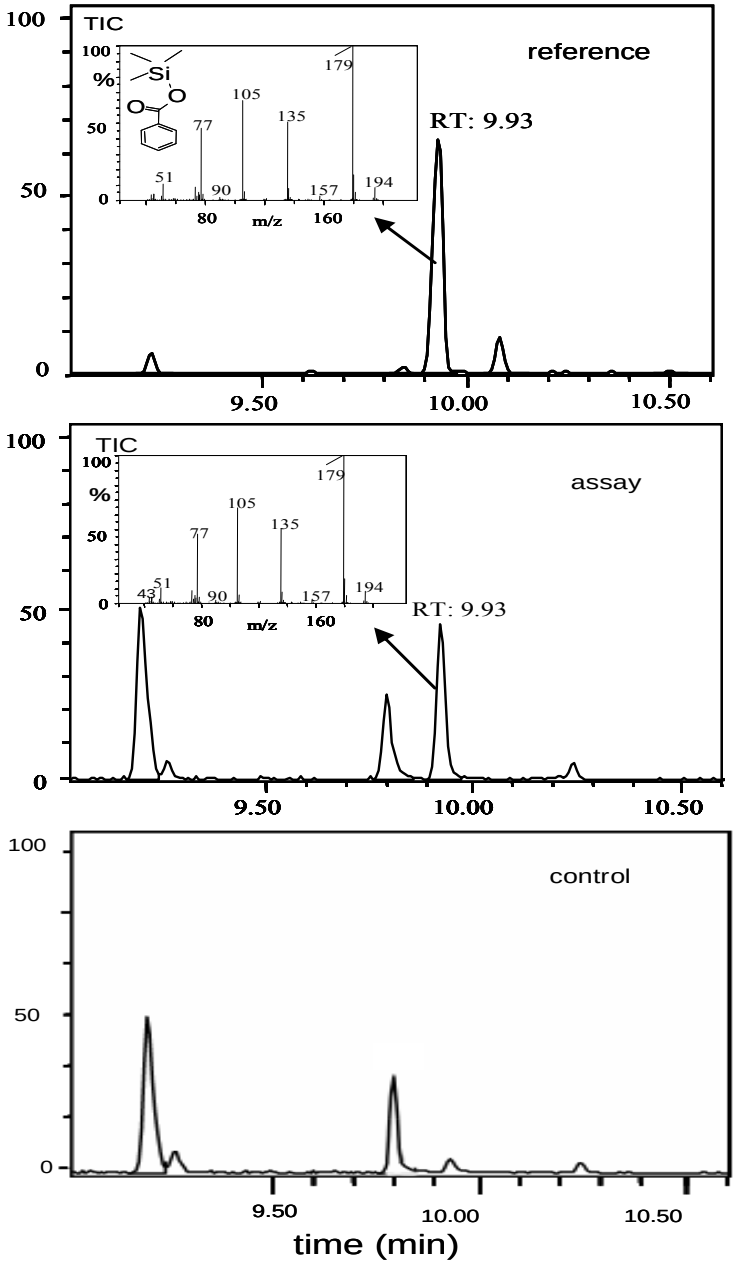


Figure 13. GC-MS analysis of benzaldehyde dehydrogenase assays. The enzymatic product and the authentic reference compound (benzoic acid) were derivatized with MSTFA. The control incubation contained heat-denatured protein. Inserts show the mass spectra of enzymatically formed and authentic benzoic acid (tri-methylsilyl ester-derivative; RT: 9.93).

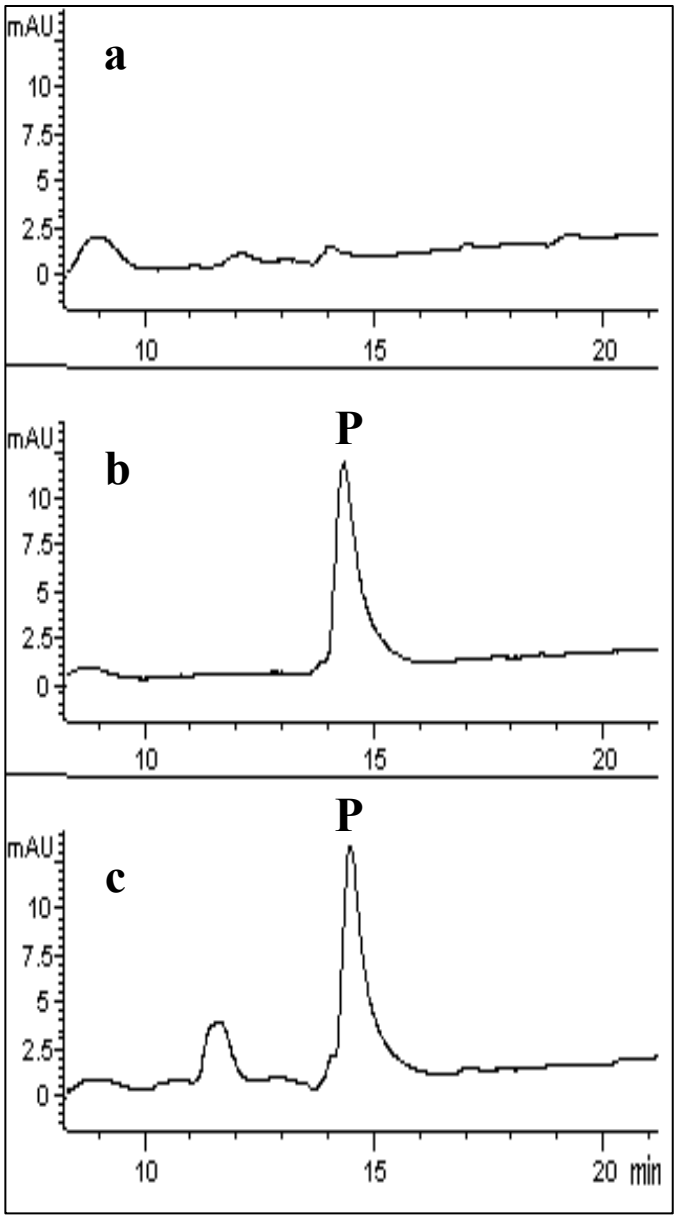


Figure 14. HPLC analysis of biphenyl synthase assays: (a) incubation with boiled protein, (b) standard incubation with 2 µg recombinant protein (positive control), (c) standard incubation with desalted crude protein extract. P, enzymatic product (3,5-dihydroxybiphenyl). Detection wavelength, 281 nm.

1.1 Changes in BD and BIS activities after elicitor treatment

Chitosan-treated cell cultures were harvested at different times after the onset of elicitation (3-24 h). Cell-free extracts were prepared and used for the determination of BD activity. Up to 12 h post-elicitation, there was a low basal level of enzyme activity (around 10 pkat/mg protein) [Figure 15 a]. Thereafter, the enzyme activity rapidly increased and reached a maximum at 17 h after the onset of treatment (36.71 ± 0.87 pkat/mg protein). At 24 h post-elicitation, BD activity had nearly decreased to the basal level. Non-elicited control cells showed no or little changes in BD activity over the period studied. The maximum enzyme activity in treated cells (17 h post-elicitation) was 3.5 times that of control cells.

A similar transient induction by chitosan treatment was observed for BIS activity [Figure 15 b]. However, neither control cells nor treated cells at early time points post-elicitation contained a basal level of BIS activity. This finding agrees with a recently published RNA blot analysis (Liu et al., 2007).

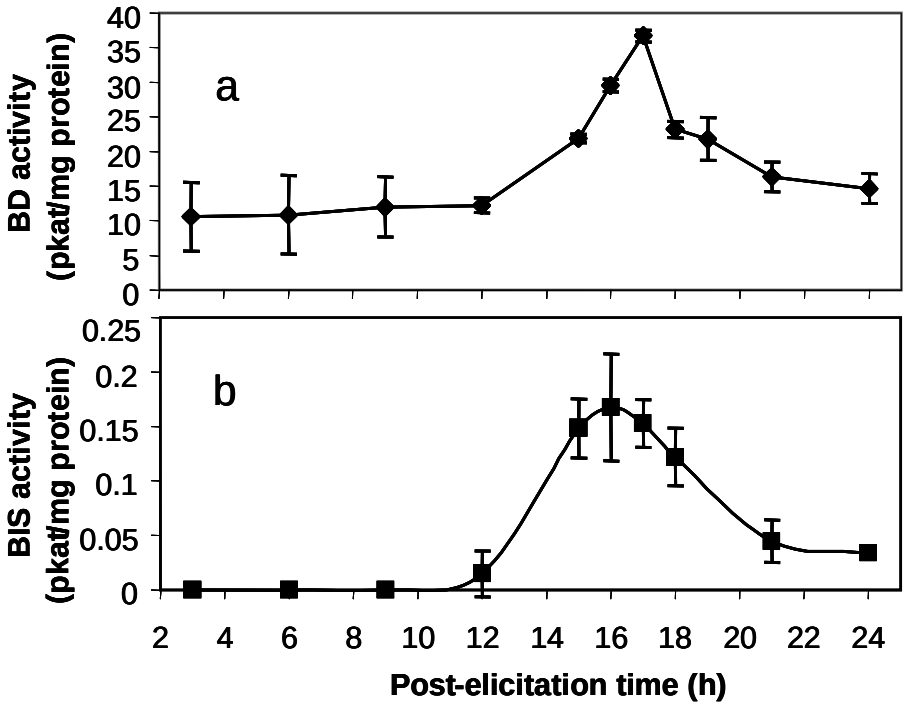


Figure 15. Changes in benzaldehyde dehydrogenase (BD; a) and biphenyl synthase (BIS; b) activities in 5-day-old *S. aucuparia* cell cultures after treatment with chitosan (25 mg/L). Each value is the mean ± SD from three independent experiments.

1.2 Biochemical characterization of BD

1.2.1 Determination of pH and temperature optima

The enzyme showed a pH optimum at 9.5 [Figure 16 a]. The temperature optimum for the reaction was 40°C [Figure 16 b]. An incubation temperature of 40°C and a pH value of 9.5 were used in all subsequent experiments for further characterization.

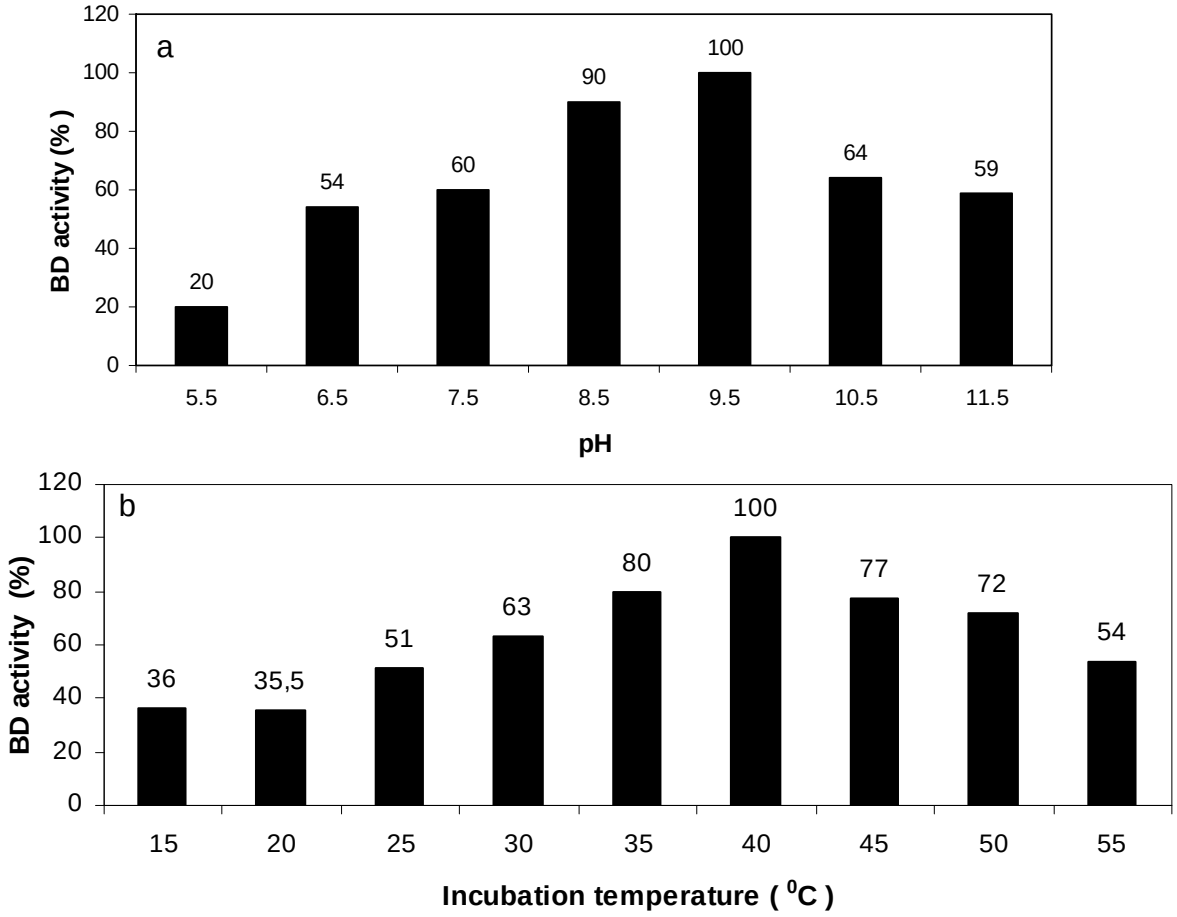


Figure 16. pH and temperature optima of BD. (a) pH optimum, (b) temperature optimum. Results are the means of triplicate experiments.

1.2.2 Linearity with protein concentration and incubation time

Benzoic acid formation was linear with protein amounts up to 130 µg in the standard assay [Figure 17 a] and with the incubation time up to 30 min [Figure 17 b]. An incubation time of 30 min and a protein concentration of 130 µg per assay were used in all subsequent experiments for enzyme characterization.

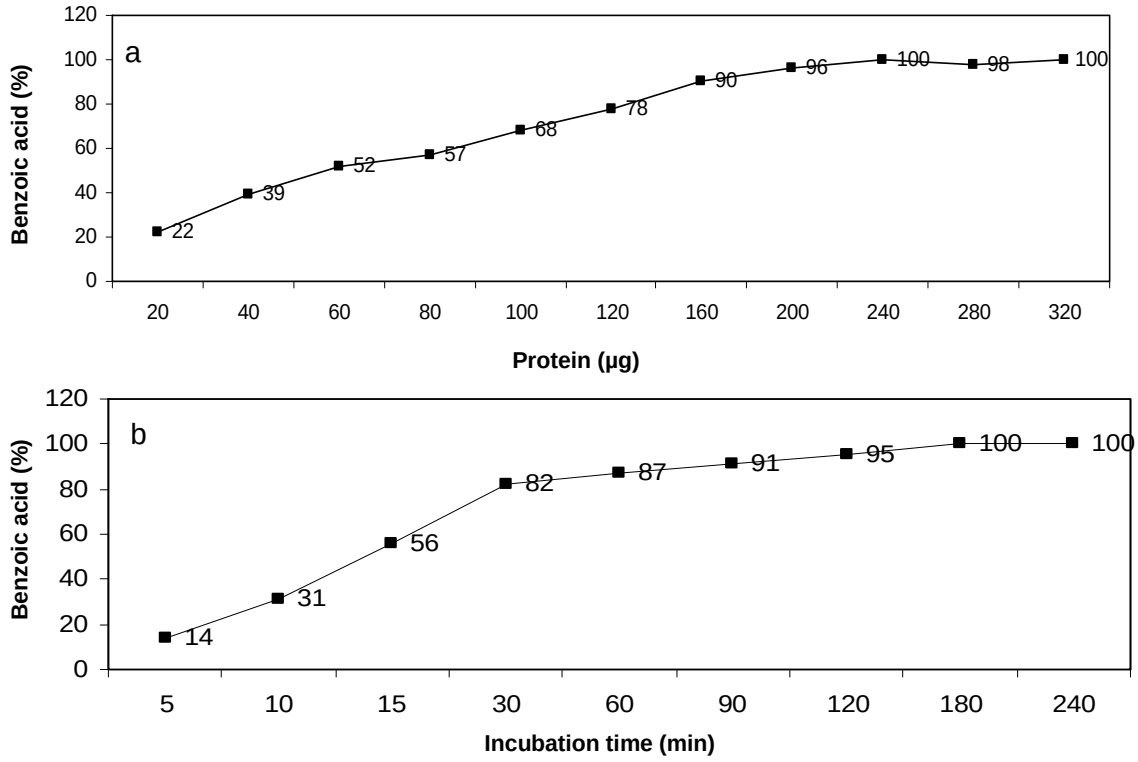


Figure 17. Dependence of BD activity on protein amount (a), and incubation time (b). Results are the means of triplicate experiments.

1.2.3 Thiol dependence

Addition of DTT increased BD activity [Figure 18] (28% increase in the presence of 50 µM DTT). However, it also stimulated the enzymatic formation of a polar unidentified side product and was therefore not included in the standard assay.

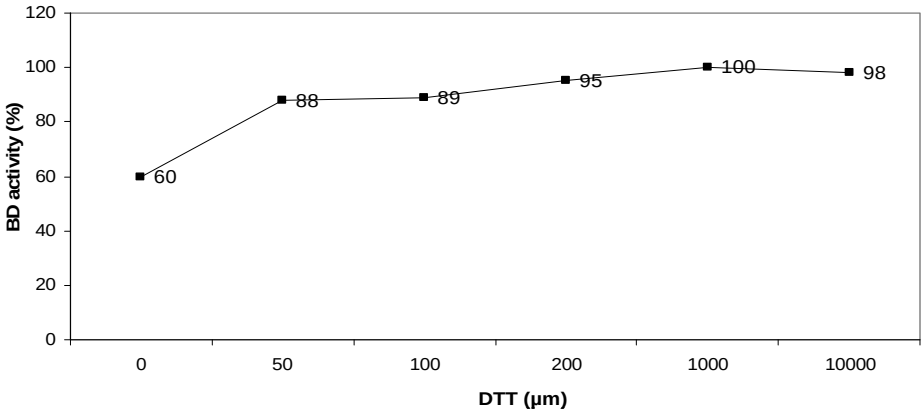


Figure 18. Effect of thiol reagent on BD activity.

1.2.4 Effect of divalent metal cations

Various divalent metal cations were tested at 1 mM for their possible inhibitory or stimulatory effect on BD activity. None of the ions tested influenced the BD activity. The greatest inhibitory effects, 70 and 77% reduction in activity, were observed in the presence of Zn^{++} and Cu^{++}, respectively [Figure 19].

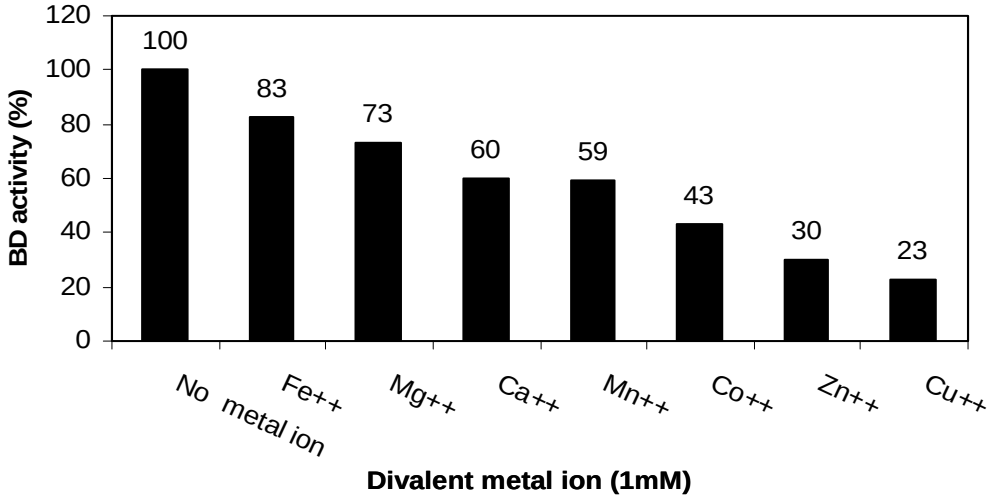


Figure 19. Inhibition of BD activity by divalent metal cations (1mM). Standard assay lacking a metal ion considered as control.

1.2.5 Effect of cofactors

NAD^+ was the preferred cofactor (100%), but $NADP^+$ was also relatively efficient (86%). Markedly lower catalytic activities were found with FAD (37%) and FMN (31%) [Figure 20].

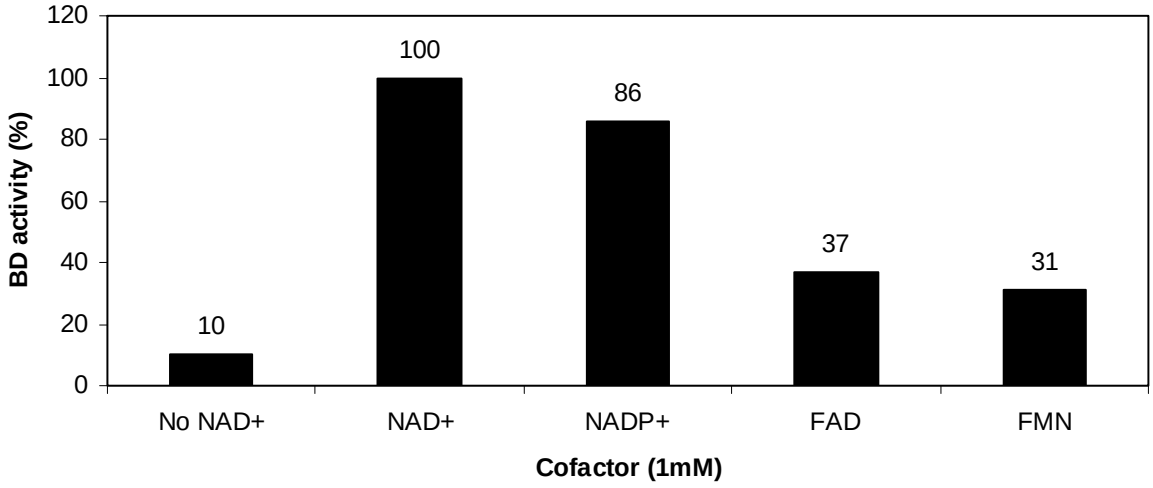


Figure 20. Effect of cofactors (1mM) on BD activity.

1.2.6 Substrate specificity

Six different aromatic aldehydes were tested as potential substrates at saturating concentrations of 0.5 mM [Figure 21]. The preferred substrate for BD was benzaldehyde (100 %) followed by cinnamaldehyde (29.3%). Hydroxylated and methoxylated derivatives were poor substrates.

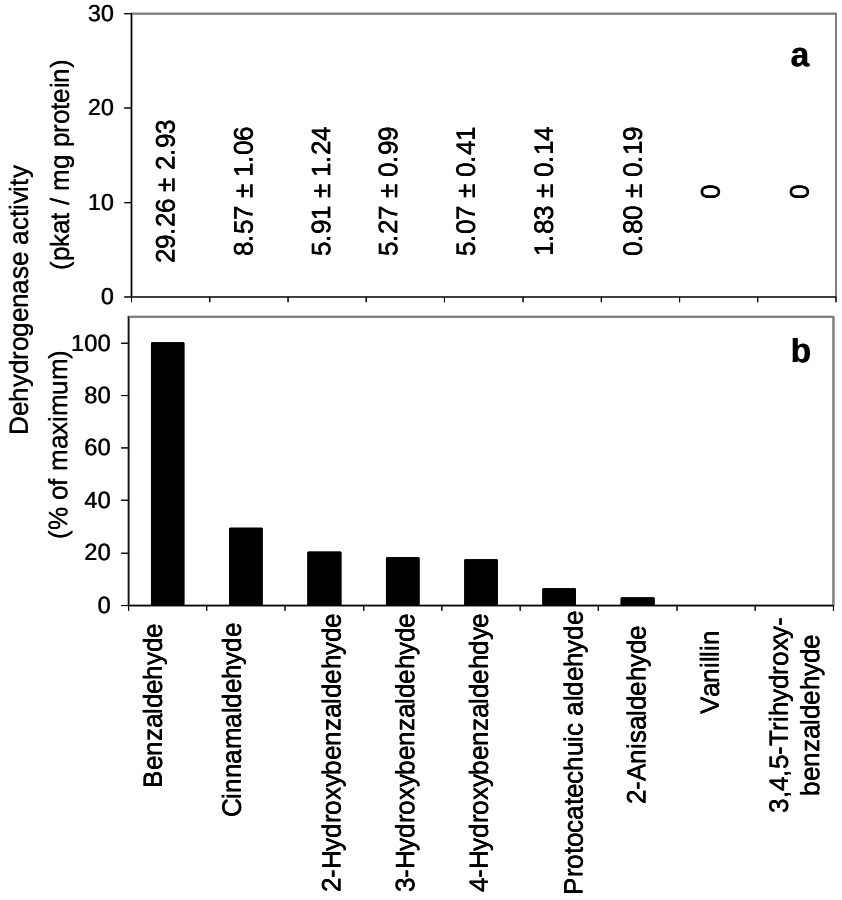


Figure 21. Substrate specificity of BD. Data are presented as specific enzyme activities (a) and percentage values (b). The results are means ± SD from three independent experiments.

1.2.7 Enzyme stability

When cell-free extracts were frozen in liquid nitrogen and stored at -20°C for 24 h or kept at 4°C for 24 h, BD activity decreased about 27 and 36%, respectively, from its initial activity [Figure 22].

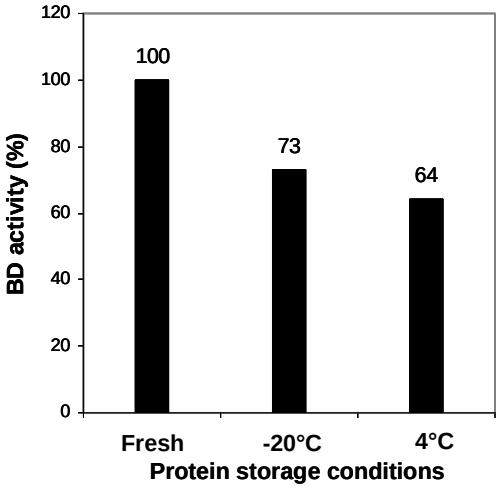


Figure 22. Stability of BD.

1.2.8 Kinetic parameters of BD

The increases in reaction rate with increasing concentrations of benzaldehyde and NAD^+ obeyed Michaelis-Menten kinetics. The apparent K_m values, as determined from Hanes plots, were 49 µM for benzaldehyde and 67 µM for NAD^+ [Figure 23].

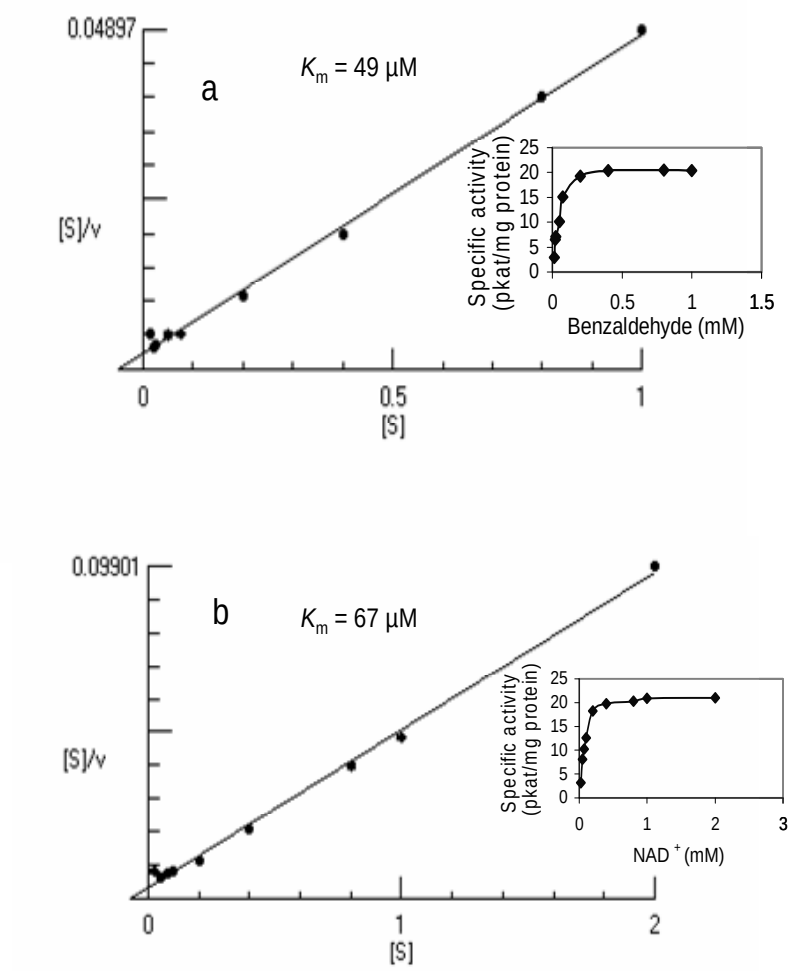


Figure 23. Determination of the apparent K_m values for benzaldehyde (a) and NAD^+ (b) from Hanes plots. Inserts show the Michaelis-Menten kinetics.

2 Cloning of cDNA encoding CoA ligase

2.1 An overview

The presence of several conserved peptide motifs in all available 4CL amino acid sequences has been repeatedly observed by computer assisted sequence alignments (Becker-André et al., 1991; Allina et al., 1998; Hu et al., 1998; Ehlting et al., 1999). The Box I motif, SSGTTGLPKGV, is not only absolutely conserved in 4CLs, but highly similar motifs are also found in luciferases, acetyl-CoA synthetases, long-chain fatty acyl-CoA synthetases and non-ribosomal polypeptide synthetases (Fulda et al., 1994). The presence of this putative nucleotide binding motif has been used as one important criterion to establish the superfamily of adenylate-forming enzymes. The Box II motif, GEICIRG, is absolutely conserved in all 4CLs. In our group, degenerate primers targeting several conserved regions within the *4CL* genes were designed. Forward and reverse primers were derived from the first putative AMP-binding domain (Box I) and the second AMP-binding domain (Box II) using 4CL sequences accessible from the National Center for Biotechnology Information (NCBI). Two full-lengths cDNAs encoding 4CL2 (Scharnhop, 2008) and 4CL3 (Ramadan, 2006) as well as a cDNA fragment encoding 4CL1 (Ramadan, 2006; Scharnhop, 2008) were cloned from elicitor-treated *S. aucuparia* cell cultures. Based on sequence data of the cDNA fragment (4CL1), trials have been done here to get its 5'-end and to functionally express the 4CL1 coding region in *E. coli*. The three Sa4CLs show high homology to the corresponding isoforms from *Rubus idaeus*, also belonging to the family Rosaceae. It was postulated that use of degenerate primers at AMP-binding domains must lead to cloning of 4CL cDNA. As benzoate and cinnamate-CoA ligases (BZL, CNL) were assumed to be related to 4CL, the working strategy was to clone, in addition, BZL and CNL cDNAs, if sufficiently evolutionarily related.

2.2 Attempts to clone the 4CL1 5'-end using total RNA

According to the 4CL1 sequence data obtained from Scharnhop (2008), reverse gene specific primers were designed (II.5.2.1). Total RNA [Figure 24] was extracted from 100 mg of cell suspension culture as described in II.3.2. For reverse transcription (II.3.4), 1µg total RNA was used. The obtained cDNA was examined for its quality by checking the successful expression of the three Sa4CL isoforms. The forward and reverse gene specific primer pairs (II.1.5.2.7) were used. With standard PCR programme and annealing temperature of 55°C, the primers were designed to yield a PCR product of 898, 897, and 896 bp for Sa4CL1, Sa4CL2 and Sa4CL3, respectively [Figure 25].

The above cDNA pool containing the 4CL sequences was used as a template for 5`-RACE (II.3.4.1) [Figure 26] together with reverse gene-specific primers in order to identify the 5'-end of the 4CL1 clone. However, neither changes in the annealing temperature (TA) nor the other PCR conditions led to isolation of the right PCR product confirming the 4CL1 5'-end.

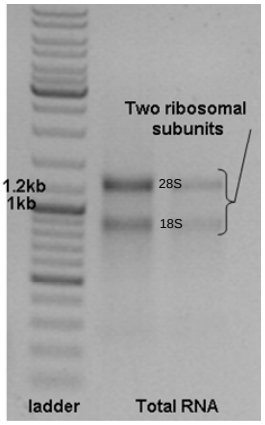


Figure 24. Gel electrophoresis of total RNA extracted from *S. aucuparia* cell suspension.

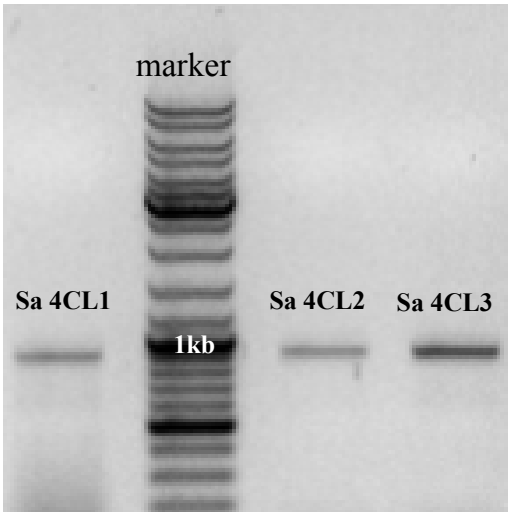


Figure 25. RT-PCR amplification of transcript fragments encoding the Sa4CL isoenzymes 9h after elicitor treatment. Template dilution, 1:10.

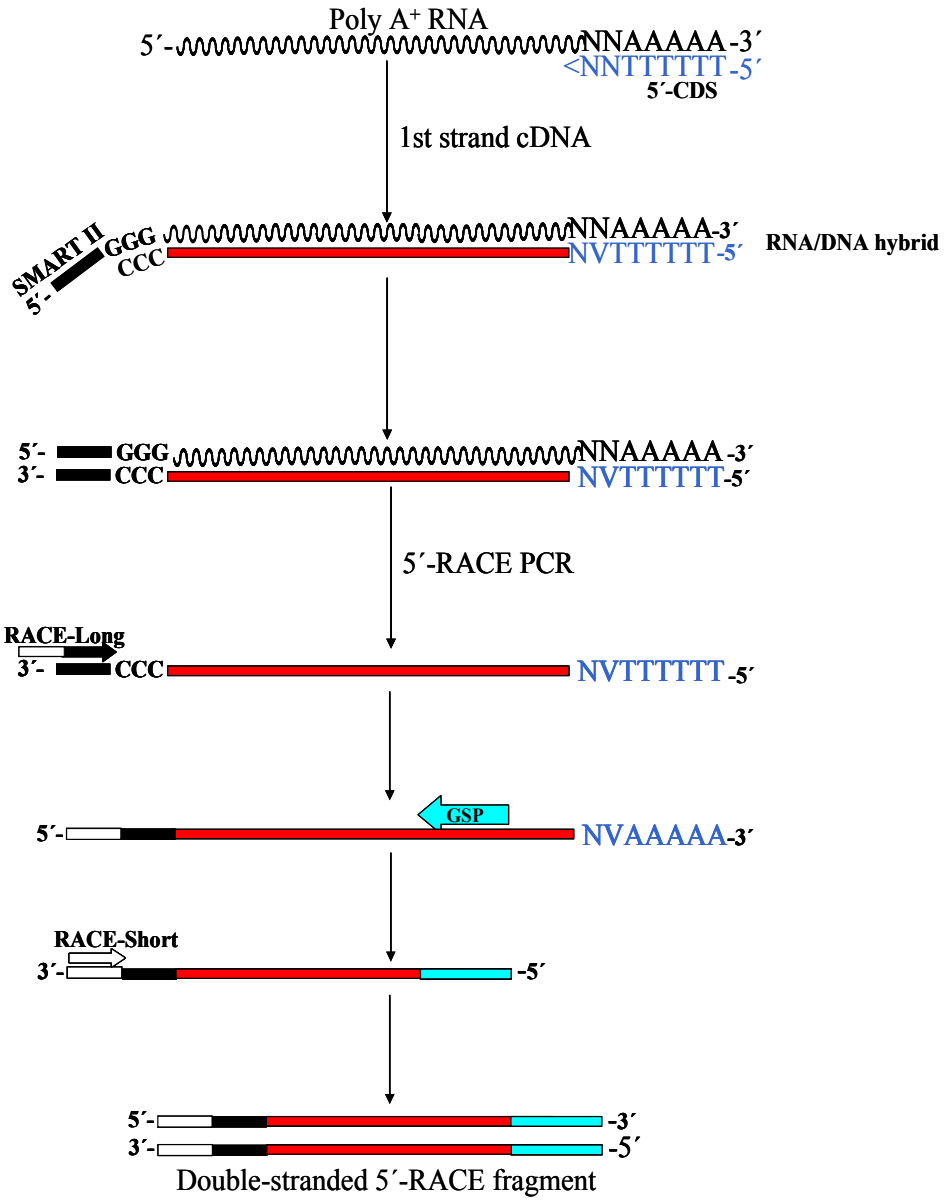


Figure 26. Detailed representation of the 5′-RACE reactions.

2.3 Cloning of 5'-ends of CoA-ligase sequences from genomic DNA

In the RACE protocol, 3'RACE is easier to give success compared to 5'RACE, which is due to the presence of a polyA tail at the 3'end of mRNA. Hence, new strategies of 5'RACE have been developed (Schramm et al., 2000; Scotto-Lavino et al., 2006). However, obtaining the 5'-end of a gene of interest for generating the full-length open reading frame (ORF) is indeed a tedious task due to the many-steps approach, a low gene copy number and cautious manipulation of RNA.

Genomic DNA was isolated (II.3.2) followed by running a semi-nested PCR with TD primers (II.3.5.3). TATA-Box degenerate (TD) primers (II.1.5.2.6) were designed according to the TATA-box, which is conserved in the promoter region of genes in eukaryotic organisms [Figure 27] (Guo et al., 2010). The pGEM-T Easy vector [Figure 28] allows the direct cloning of *Taq* polymerase-amplified PCR products. The insert can be released by single digestion using *EcoR*1.

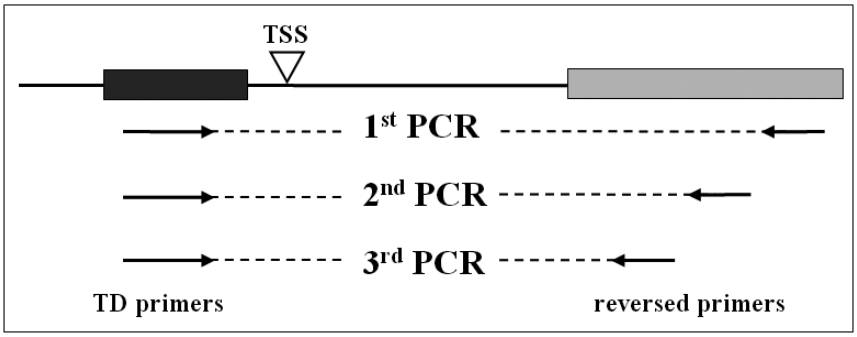


Figure 27. Scheme of semi-nested PCR with TD primers. Black, TATA-box; Grey, known region of gene of interest; TSS, transcription start site (Guo et al., 2010)

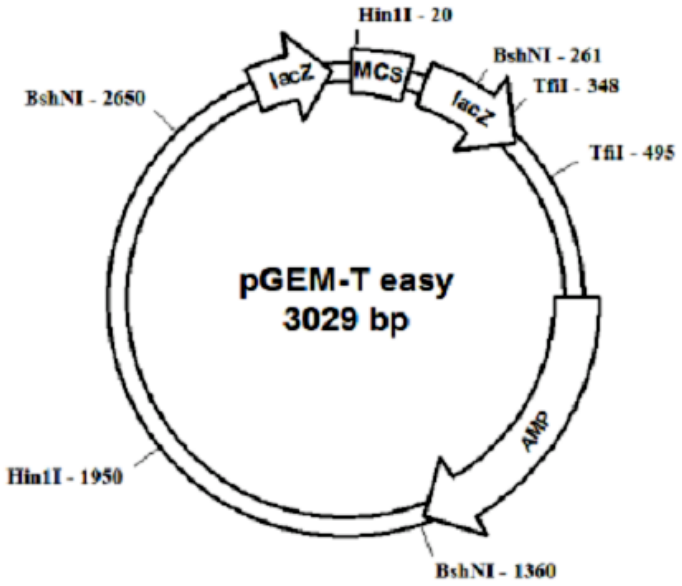


Figure 28. pGEM-T easy vector

2.3.1 Identification of a new 5'-end of a CoA-ligase sequence

The attempts to get the 5'-end of the Sa4CL1 clone resulted in isolation of an additional cDNA fragment encoding 4CL or 4CL-like [Figure 29]. The fragment was cloned by using the reverse gene-specific primer P4 (II.5.2.1) in the third amplification step of the TD-semi-nested PCR (II.3.5.3; Table 1). The PCR product was about 780 bp (483 bp is partial promoter and 5´-untranslated region (UTR) and 297 bp is coding region including the start codon). The fragment was deposited in the data bank (accession number GU938594). When compared to the corresponding fragment of the 4CL1 cDNA, 81% nucleotide sequence identity was found. The new fragment was referred to as 4CL4 sequence. The new 5'-end was used to derive forward gene-specific primers (II.1.5.2.2) for 3'-RACE (III.3).

Table 1. TD-semi-nested PCR primers

Reverse gene-specific primers (II.5.2.1) used in semi-nested PCR to get the 5´-end of a new *4CL* gene	Amplifications
3	1st
3	2nd
4	3rd

Product size = 780 bp

5´-CTTGTCGTAGTAGCACGCTAAGCAGGTCTAGACACACTTTCTCAATTTCCAAC
GCTTAGGAAGTAATCGGGTTGTTAAACAGCAGCCTAACACCTAGGCTAGGCC
TAGGCGGGAATTTTTAGAACAGTGTACAGAAGGCATCCAATGTGGCACGCAC
ATACTTAAAAATAGACACGTTTATATCAGTAAATTTTTCTTCTTCCTCCAGCT
GATCATAGTTGGAAAGTAAAAGAATGCAGAACAACTGCATATACTGGTAGTG
ATTACAGCTTCCCATTGTGCGAGACAACCTCAAACGCTAGACACTAATCGAG
CAATGAGCTAGGGCCTAGTCGCCTCCCCTCCCCCAACTAACCcACAACTCTTG
TTGTTTATACTATATGAAGATCCCTATGGGCCTATAAAAGCTCCAATTTTGCT
CAGAACAAACCAATCATTTTGGTGTCTTTTTTCTCTTTGGATAATCCCATTTTT
ACCAAGAGATGGCTATAGAAACTACCCAAAACGACATCGTCTACCGACCC
AAAATCCCCGACATCCCTATCTCAAAACACCTCCCTCTCCATTCCTACTG
CCTTCGCAACAAAAACCACCCGAGCTCCAAACCCGCATCATCAACGACG
CCACTAGAGACATCTACACATACACTGATGTGGAACTCAACGCGCTCAG
AGTTGCCTTAGGGCTAACAAGCTCGGAATCCAACAGGGTGACGTTATCA
TACTATTCCTCCCCAACTCGCTGAAGTTTATCTTCTCCTTCCTCGAGCCT
CCTTT-3´

Figure 29. Upstream nucleotide sequence of a new *Sa4CL* gene (*4CL4*). Black, 5'-untranslated region; red, coding region; underlined, start codon.

2.3.2 Identification of the 5´-end of the *Sa4CL1* gene

Using genomic DNA as a template and TD-degenerate primers together with reverse gene-specific primers (II.5.2.1; Table 2), a PCR product of 624 bp was obtained through a semi-nested PCR protocol (360 bp is partial promoter and 5´UTR and 264 bp is coding region including start codon).

Table 2. TD-semi-nested PCR primers

Reverse gene specific primers (II.5.2.1) used in semi-nested PCR to get the 5´-end of the *Sa4CL1* gene	Amplifications
2	1st
2	2nd
5	3rd

Product size = 624 bp

2.4 Construction of 4CL1/pRSET B plasmid

By using softwares for sequence analysis, both previously cloned fragments of the 4CL1 cDNA (Ramadan, 2006; Scharnhop, 2008) and its 5'-end (III.2.3.2, the present work) were combined to construct the complete hypothetical full-length cDNA. Nucleotide sequence information obtained from this construct allowed the design of a new pair of GSPs (II.1.5.2.3) to amplify the whole ORF from the start codon ATG until the stop codon, TGA. Platinum *Pfx* polymerase was used to reamplify the whole ORF from reversely transcribed elicitor-induced RNA in a standard PCR at an annealing temperature of 55.5°C (II.3.5). The primers were designed so that the forward primer, Sa.Exp X1-5' *Nhe*I, contained an *Nhe*I restriction site and integrated the start codon ATG, whereas the reverse primer, Sa.Exp.X1-3' *Kpn*I, contained a *Kpn* I restriction site directly behind the stop codon TGA (II.3.8.2). The restriction sites were introduced into the sequence to allow cloning of the full-length 4CL1 cDNA in the expression vector pRSET B [Figure 30]. The constructed plasmid was sequenced to ensure the presence of the right insert (4CL1 coding region) and no frame shifts were detected. Heterologous expression of Sa4CL1 as a His_6-tagged protein failed. Despite variation in the expression conditions, no soluble protein was obtained. However, expression as a GST-fusion protein by using the pGEX vector resulted in formation of soluble Sa4CL1 protein. (III.2.5, III.2.6).

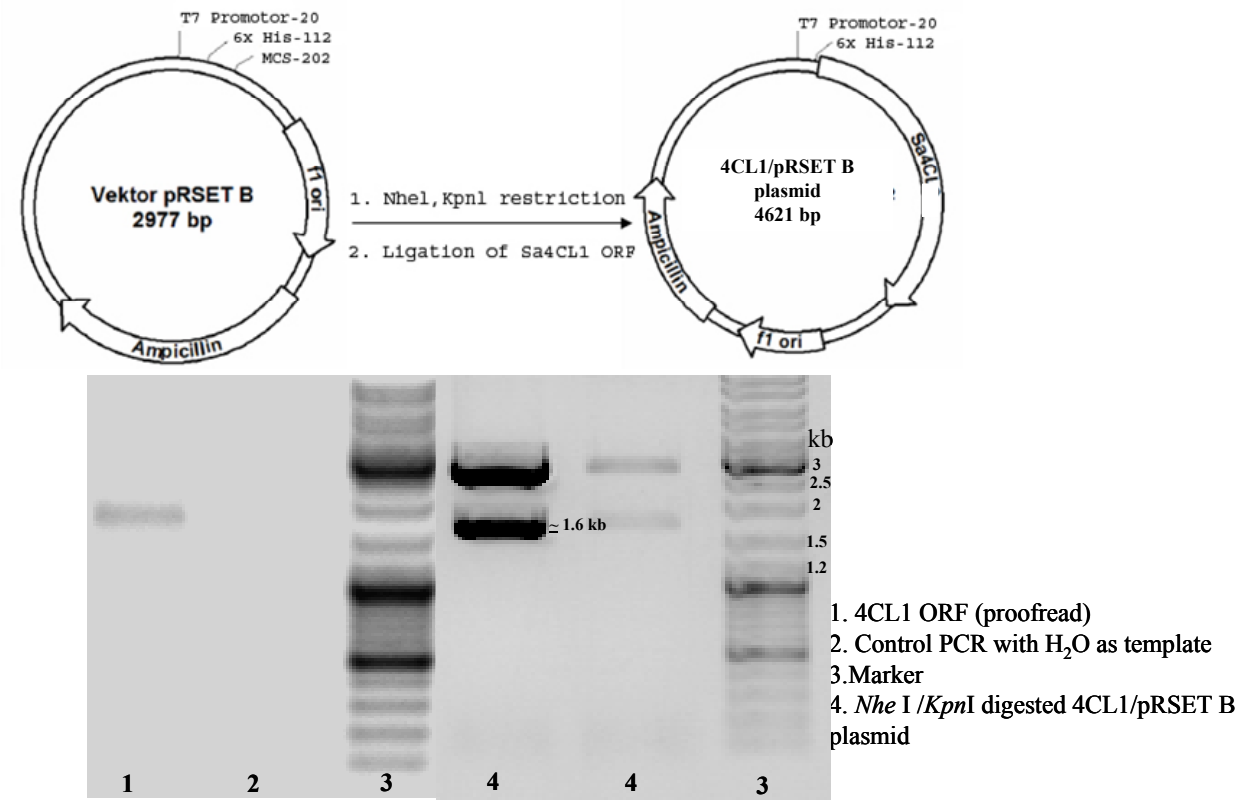


Figure 30. Plasmid construction for 4CL1 overexpression using pRSETB vector.

2.5 Construction of 4CL1/pGEX-G plasmid

Since overexpression of 4CL1 using pRSET B failed, an alternative strategy was to exchange the expression vector. The ORF was reamplified using other overexpression primers (II.1.5.2.3) that can fit with pGEX-G expression vector [Figure 31]. The primers were designed so that the forward primer, Exp 4CL-X1-ATG, did not include any restriction site and started from the start codon, whereas the reverse primer, Sa Exp X1-3' *Kpn*I, was the same as the pRSET B reverse overexpression primer. Platinum *Pfx* polymerase was used to reamplify the whole ORF. The pGEX-G expression vector was designed for any ORF to be ligated to the 3'-end of the vector without further manipulation (Görlach and Schmid, 1996; Smith and Johnson, 1988). The AGG triplet which codes for an Arg represents also the 5'-half of the AGGCCT restriction site for *Stu*I. *Stu*I cleaves in the middle of the recognition sequence and generates blunt ends at the 3'-end of the vector [Figure 31b]. The constructed pGEX-G plasmid was digested with *Xho*I (cut at position 112 of the insert) and *EcoR*I (cut in the vector after the stop codon) to release the insert [Figure 32]. The nucleotide sequence of the Sa4CL1 full-length cDNA is found in the appendix. The resulting sequence confirmed the presence of the correct insert with no frame shifts. The constructed plasmid pGEX-G/4CL1 could be successfully overexpressed in *E. coli* (III. 2.6). The 1644 bp 4CL1-ORF (accession number GU938595) encoded a 59.5 kDa protein that consisted of 547 amino acids and had an isoelectric point (p*I*) at pH value 5.5. This p*I* of Sa4CL1 was calculated from the obtained gene sequence data. The Sa4CL1 amino acid sequence exhibited 71 and 63.5% identities with the Sa4CL2 and Sa4CL3 sequences, respectively, and it showed 84% identity with *Rubus idaeus* Ri4CL1.

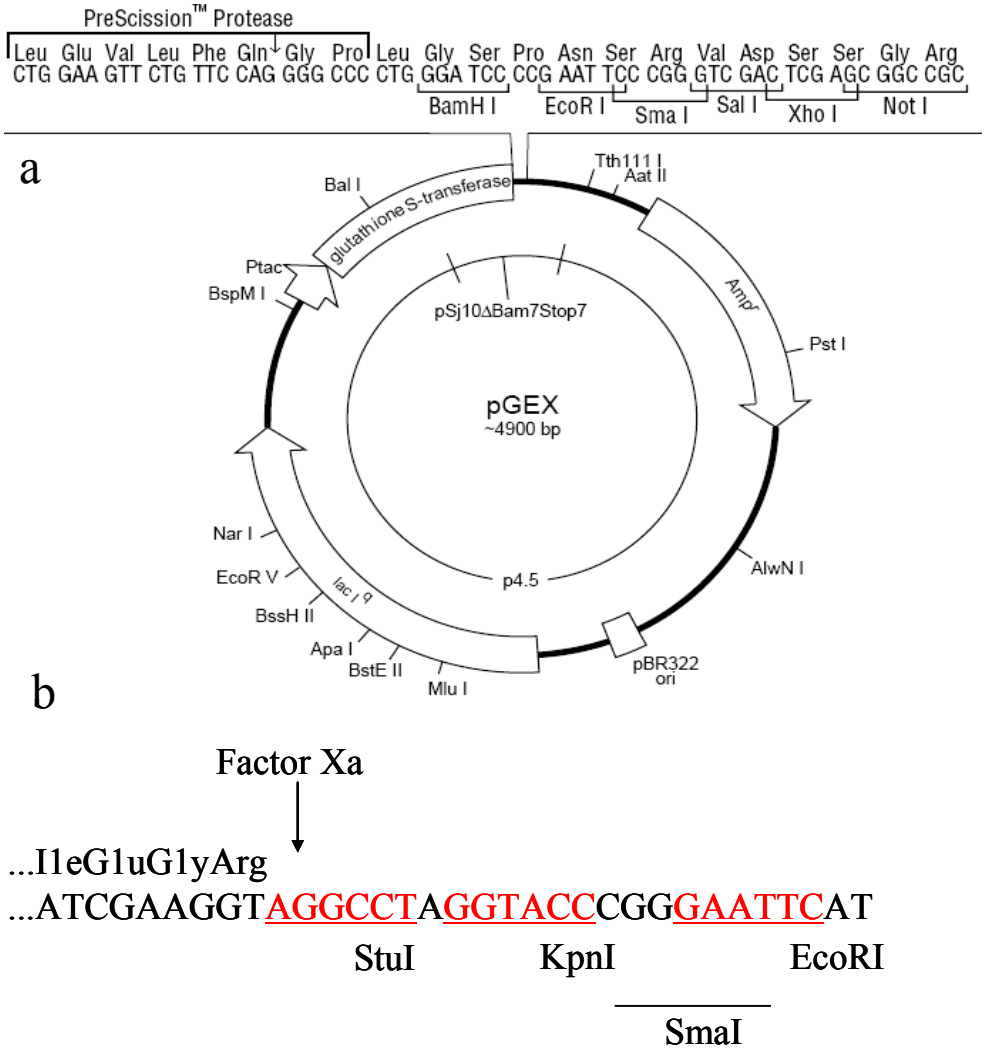


Figure 31. (a) pGEX expression vector, (b) pGEX-G vector; the Arg-encoding triplet AGG in the factor Xa recognition sequence I1e-G1u-G1y-Arg can be used to create a *Stu*I restriction site (AGGCCT) in vectors such as pGEX designed for the expression of fusion proteins (Smith and Johnson, 1988).

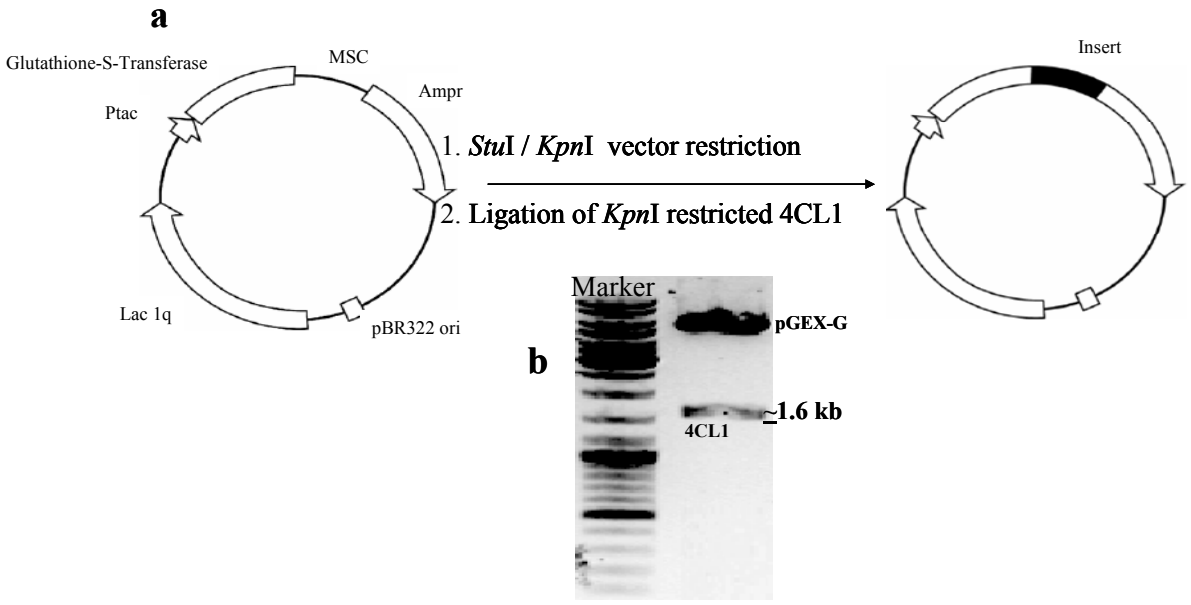


Figure 32. (a) Plasmid construction for 4CL1 overexpression using pGEX-G vector, (b) *Xho*I /*EcoR* I digested 4CL1/pGEX-G plasmid.

2.6 Expression of 4CL1 in *Escherichia coli*

The plasmid was successfully overexpressed in the *E. coli* strain BL21. Following protein extraction and purification (II.3.13), the efficiency of overexpression was examined by SDS-PAGE [Figure 33]. The yield of pure protein, as determined by the Bradford method, was around 8 mg/L bacterial culture. Purified protein could be stored in the presence of 20% glycerol at -20°C for 2 weeks without appreciable loss of activity.

Sa4CL1 was heterologously expressed as GST-fusion protein with a molecular mass of around 85 kDa including the GST-tag. The molecular mass value predicted from the amino acid sequence is 59.5 kDa after subtracting the 26 kDa GST-tag.

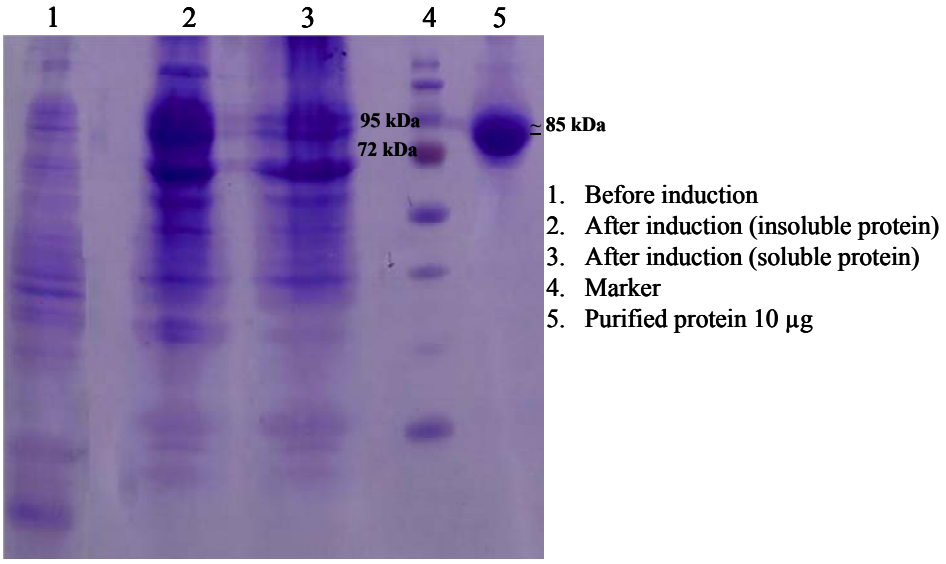


Figure 33. Purification of 4CL1 monitored by SDS-PAGE. Protein bands were visualized by staining with Coomassie blue.

2.7 Preliminary study of CoA ligase activity

The overexpressed crude recombinant protein (II.3.12) was tested for its catalytic activity prior to purification. Total protein was extracted by sonification of the harvested bacterial pellets. The extracted soluble protein (50 µl) was used in spectrophotometric assays (II.2.12) to qualitatively determine the CoA ligase activity. CoA ligase assays were performed using a series of substrates (*p*-coumaric, caffeic, ferulic, cinnamic, and sinapic acids). The crude recombinant protein exhibited CoA ligase activity with most substrates. The reaction was linear with time up to 5 min followed by decrease in the reaction rate as observed from the recorded absorbance. It appeared that poor yield and limited linearity with the crude enzyme preparations may reflect a balance between CoA ligase and thioesterase activity. Crude enzyme extracts might also contain other ATP-utilizing enzymes, which compete for the limited amount of ATP. It has been reported before that soluble protein fraction of cell-free extracts from *E. coli* BL-21 expressing tobacco 4CL showed high thioesterase activity, which resulted in a low yield of CoA ester (not exceeding 40%; Beuerle and Pichersky, 2002).

The preferred substrate was *p*-coumaric acid followed by caffeic and ferulic acids. Cinnamic acid was a poorly accepted substrate. No CoA ligase activity was detected with sinapic acid.

2.8 Benzoate-CoA ligase activity

The purified recombinant protein (5 µg) was used for benzoate-CoA ligase assays (II.2.13). The substrate used was radiolabelled [7-^{14}C] benzoic acid. Radioactivity, measured in cpm (count/minute), was detected by liquid scintillation in each assay. After 30 min of incubation at room temperature, the amount of radioactivity detected in the aqueous phase of the standard assay was similar to that observed in assays containing heat-denatured protein. This result indicated that the fusion protein had no benzoate-CoA ligase activity.

2.9 Identification of the major enzymatic product

p-Coumaroyl-CoA formed by affinity-isolated recombinant protein was purified through a solid-phase extraction (SPE) LC-18 column (II.2.14). In the positive ionization mode, the mass spectrum of *p*-coumaroyl-CoA was dominated by ions that stand in good agreement with calculated molecular ions of the corresponding CoA ester [Figures 34, 35]. The sample had first been scanned for the presence of a compound with a molecular ion peak of m/z 914 $[M+H]^+$ [Figure 34 a]. In the EPI^+ experiment, the selected mass of m/z 914 was further fragmented to give the MS/MS spectrum of this compound [Figure 34 b]. A peak at m/z 768 was observed, which corresponds to the coenzyme A moiety. At m/z 407, the phosphoadenosine moiety of the CoA ester was cleaved out and the resulting fragment $[C_{20}H_{27}N_2O_5S]^+$ was detected. The stability of this fragment was indicated by its high intensity [Figure 34 b].

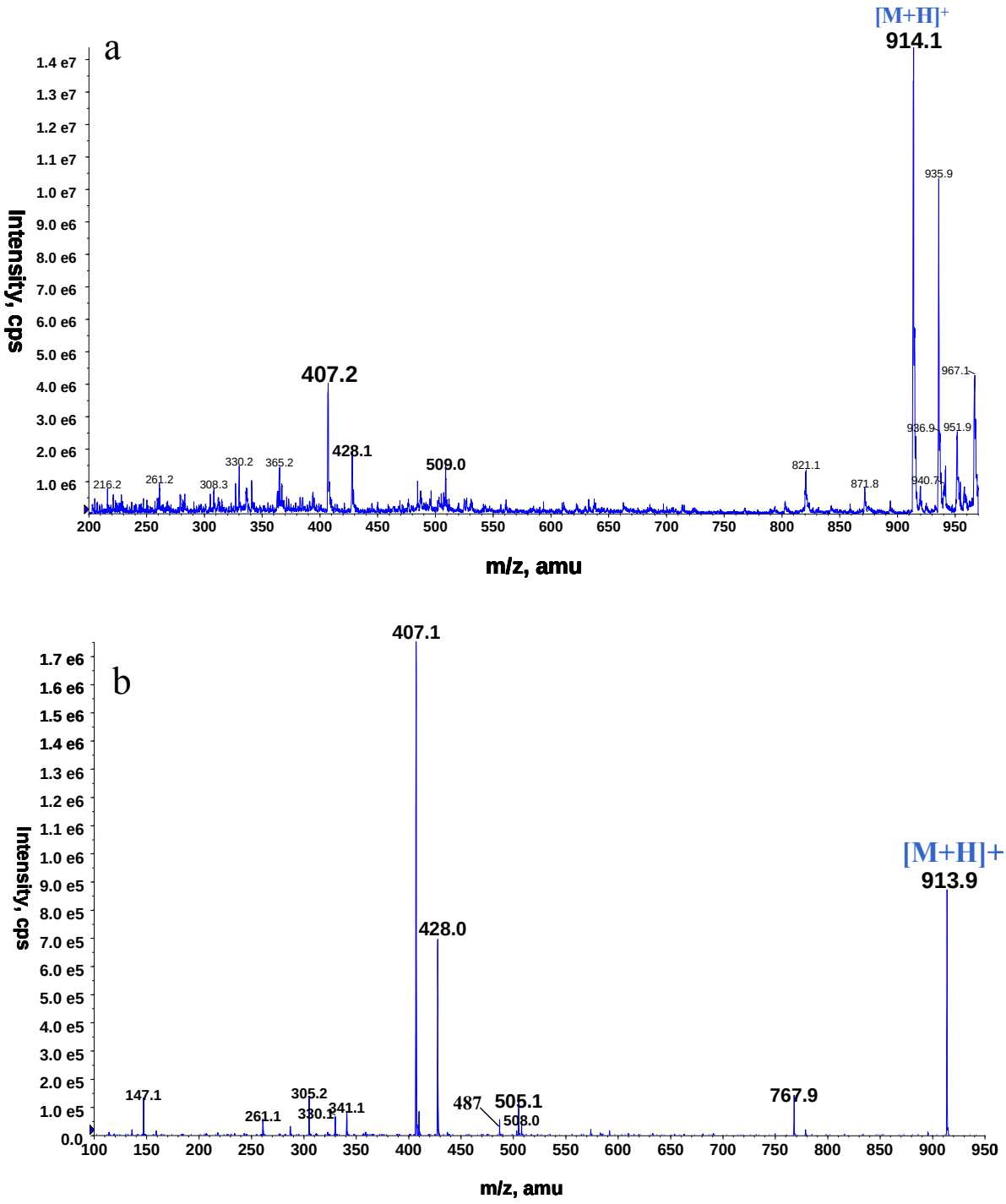
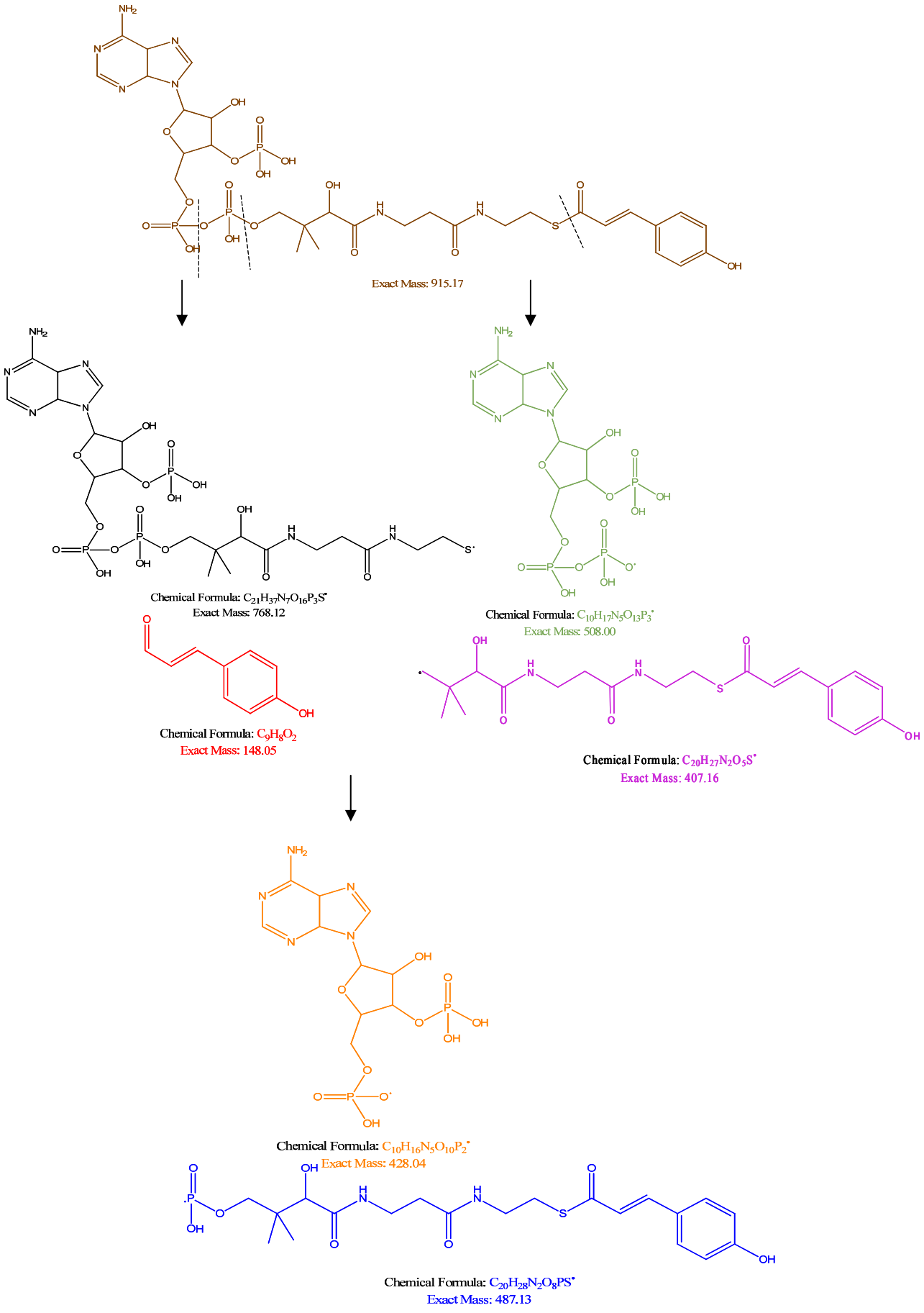


Figure 34. (a) Mass spectrum ($Q1^+$) and (b) MS/MS-spectrum (EPI^+) of purified *p*-coumaroyl-CoA.



2.10] **Figure 35.** Postulated fragmentation pattern of *p*-coumaroyl-CoA.

2.10.1 pH optimum

The pH of a solution can have several effects on the structure and activity of enzymes. For example, the pH can affect the state of ionization of acidic or basic amino acids. If this occurs, the ionic bonds that help to determine the 3-D shape of a protein can be altered. This can lead to altered protein properties or an enzyme might become inactive. Changes in pH may not only affect the shape of an enzyme but may also change the shape or charge properties of the substrate so that the substrate either cannot bind to the active site or cannot undergo catalysis. The dependence of Sa4CL1 activity on pH was studied [Figure 36]. The optimum pH was at 7.5. At this optimum pH, another series of incubations were performed at different temperatures, 25 - 50°C, to determine the optimum temperature (III.2.10.2).

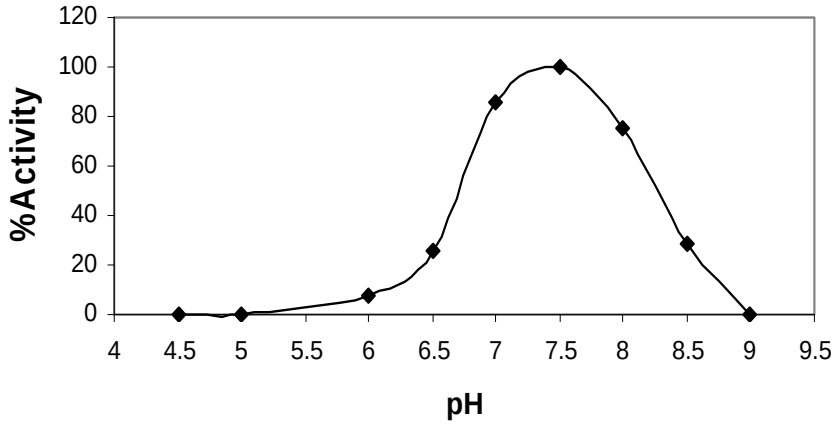


Figure 36. Dependence of Sa4CL1 activity on pH.

2.10.2 Temperature optimum

All chemical reactions respond to temperature. It is the same reason why temperature affects enzyme reactions. Enzymes are not active with temperature below 0°C. Meanwhile, at higher temperature, the velocity of the reaction increases because more energy is supplied to break the intramolecular attraction of the protein structure. In case of Sa4CL1, the velocity of the reaction increases with temperature until 40°C. Each enzyme has a temperature range in which a maximal rate of reaction is achieved. This maximum is known as the temperature optimum of the enzyme.

The temperature optimum of Sa4CL1 was between 35-40°C (37°C). At incubation temperatures above 40°C, the protein begins to denature, loses its function and becomes inactive to operate properly [Figure 37].

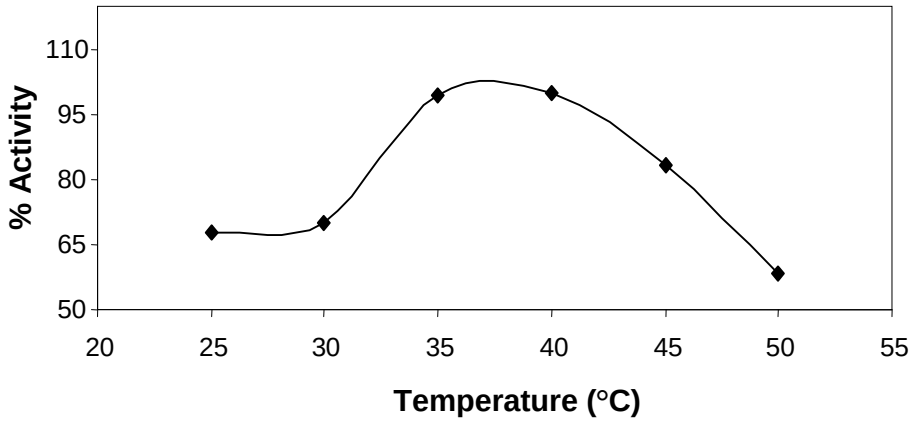


Figure 37. Dependence of Sa4CL1 activity on temperature.

2.10.3 Linearity with protein concentration

In order to study the effect of increasing enzyme concentrations on the reaction rate, care should be given to the substrate concentration not to be limiting. Any change in the amount of product formed over a specified period of time will be dependent on the level of enzyme present.

The amount of enzymatically formed *p*-coumaroyl-CoA (nmol) was determined as a function of the protein amount in the standard assay. Product fromation was linear with the protein concentration up to 10 µg in the standard assay (1 ml). Enzyme activity was inhibited at protein concentrations higher than 160 µg/ incubation [Figure 38].

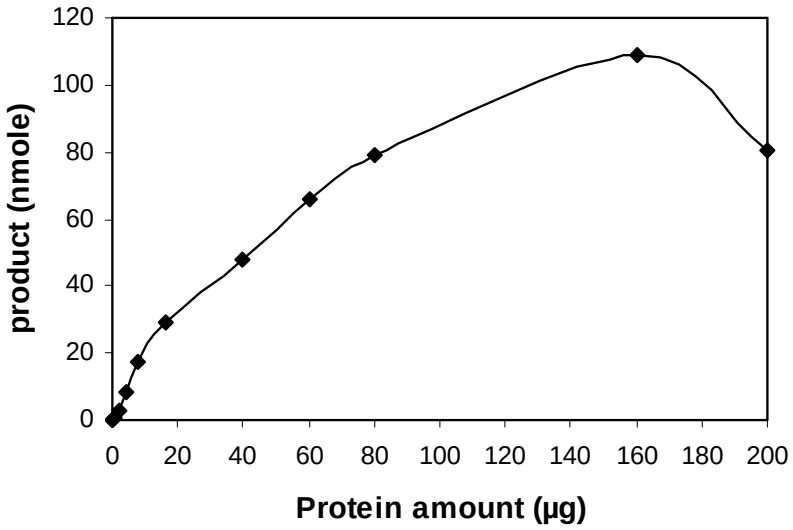


Figure 38. Dependence of Sa4CL1 activity on protein concentration.

2.10.4 Linearity with incubation time

A graph of product concentration *vs.* time follows three phases. At early time points, the rate of product accumulation increases linearly with time. For an extended period of time, the product concentration reaches a plateau and does not change with time. At later times, the substrate is depleted, so the curve starts to level off.

Enzymatically formed *p*-coumaroyl-CoA (nmole) was determined as a function of incubation time in the standard assay. Product formation was linear with incubation time up to 35 minutes [Figure 39].

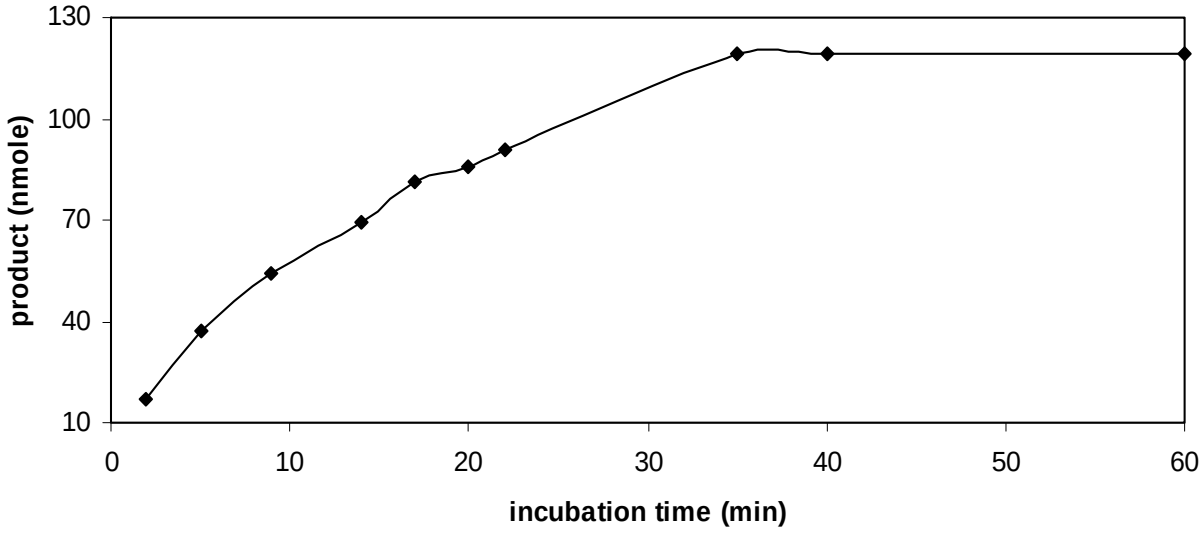


Figure 39. Dependence of Sa4CL1 activity on incubation time.

2.10.5 Effect of divalent metal ions on 4CL1 activity

Salts and ions may have many different effects of the activity of an enzyme. For example, inorganic ions may bind to some of the ionic side chains of a protein. This kind of interaction, although not always affecting the three dimensional shape of the enzyme in a substantial manner, could make it easier for a substrate molecule to locate or bind to the active site of the enzyme. Thus, the presence of the ion in optimum concentrations could alter the rate of the reaction. In some cases, an inorganic ion is an integral part of the structure of the enzyme and also participates in catalysis.

A number of divalent metal ions (Mg^{2+}, Co^{2+}, Mn^{2+}, Fe^{2+}, Ca^{2+}, Zn^{2+}, and Cu^{2+}) were tested for their effect on Sa4CL1 activity as chloride salts. Sa4CL1 activity was strictly dependent on the presence of divalent cations. The highest stimulatory effect was observed with Mg^{2+} (100% activity). Ca^{2+}, Zn^{2+}, and Cu^{2+} were least efficient [Figure 40].

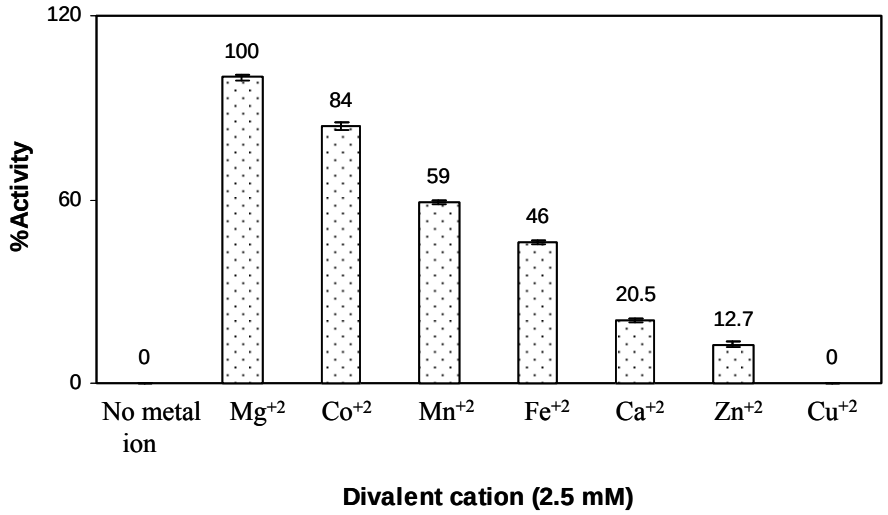


Figure 40. Effect of divalent metal ions on Sa4CL1 activity. The results are means ± SD from three independent experiments.

2.10.6 Study of substrate specificity

At the pH and temperature optima (7.5 and 37°C, respectively), spectrophotometric assays were performed using a series of substrates (*p*-coumaric, caffeic, ferulic, cinnamic, and sinapic acids) at a final concentration of 0.6 mM. The incubation time was 5 min, and 5 µg protein/ assay was used. The preferred substrate for Sa4CL1 was *p*-coumaric acid followed by caffeic acid (75% relative activity). No activity was detected with sinapic acid [Figure 41]. Radioisotopic assays were performed to test for benzoate-CoA ligase activity (III.2.8). As a positive control, benzoate-CoA ligase from the bacterium *Rhodopseudomonas palustris* was used, which efficiently catalyzed formation of benzoyl-CoA. In contrast, Sa4CL1 did not utilize benzoic acid [Figure 41].

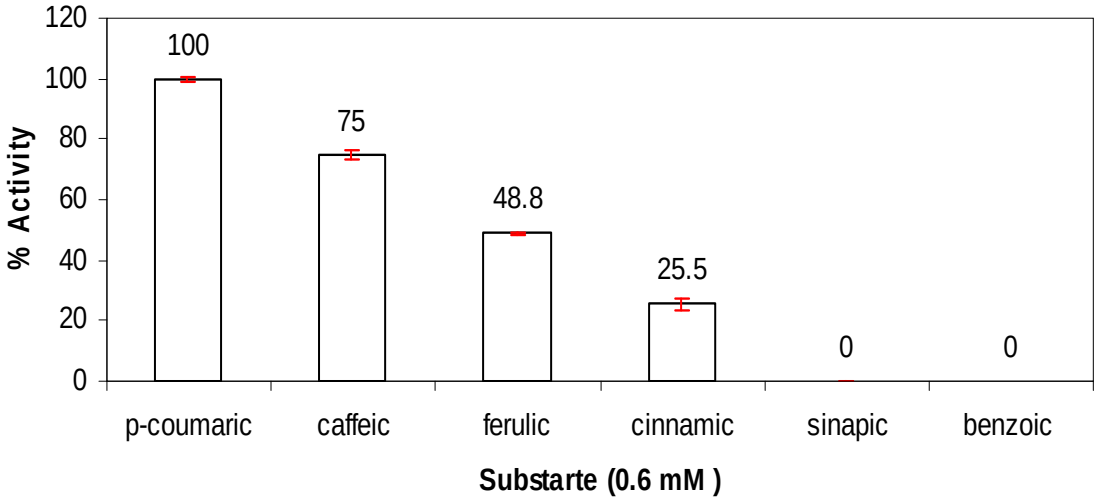


Figure 41. Substrate specificity of Sa4CL1. The results are means ± SD from three independent experiments.

2.10.7 Determination of kinetic parameters

At optimum pH 7.5, temperature 37°C, and 2 µg/assay purified protein, the kinetic properties of Sa4CL1 were determined. The increase in reaction rate with increasing concentrations of substrates (cinnamate or hydroxycinnamates, ATP and CoA) obeyed Michealis-Menten kinetics [Figures 43 a, 44, 46]. The Michaelis constant (K_m) allows for measuring the affinity of an enzyme for its substrate. A low K_m indicates that the enzyme requires only a small amount of substrate to become saturated. Hence, the maximum velocity is reached at relatively low substrate concentrations. The Lineweaver-Burk plot is a linear representation of the Michaelis-Menten equation and used to determine the kinetic parameters of an enzyme. It is a plot of the reciprocal of the enzymatic reaction velocity (1/v) versus the reciprocal of the substrate concentration (1/[S]) [Figure 42].

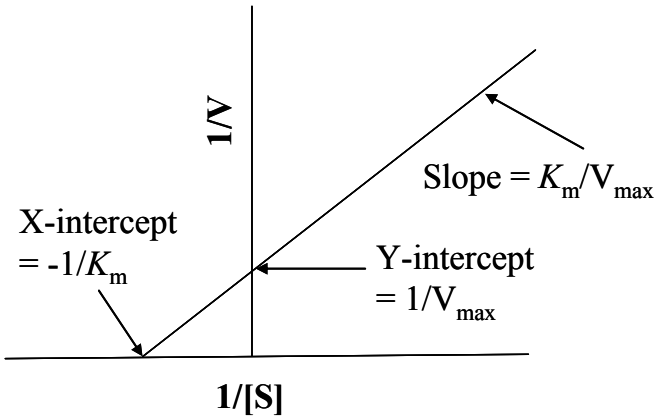


Figure 42. Lineweaver-Burk plot.

The K_{cat} value (the turnover number) is the number of substrate molecules that each active site converts to product molecules per unit time, in which the enzyme is working at maximum efficiency. It is a measure of V_{max} / enzyme concentration. The ratio K_{cat} / K_m defines a measure of the catalytic efficiency of an enzyme-substrate pair.

The K_m values of Sa4CL1, as determined from Lineweaver-Burk plots, were 7.3 µm for *p*-coumaric acid [Figure 43], 65.5 µm for ATP and 9.3 µm for CoA [Figure 44, 45]. The kinetic parameters are summarized in Table 3.

Table 3. Kinetic parameters of Sa4CL1

Substrate	K_m [µm]	V_{max} [nkat/mg]	K_{cat} [sec^{-1}]	K_{cat}/K_m [$\mu m^{-1}\ sec^{-1}$]
p-Coumaric acid	7.3	16.0	0.96	0.13
Caffeic acid	6.8	13.0	0.78	0.12
Ferulic acid	5.4	8.0	0.49	0.09
Cinnamic acid	21.0	4.5	0.26	0.01

All data are mean values from three independent experiments using affinity-purified protein. Sinapate was not accepted by 4CL1.

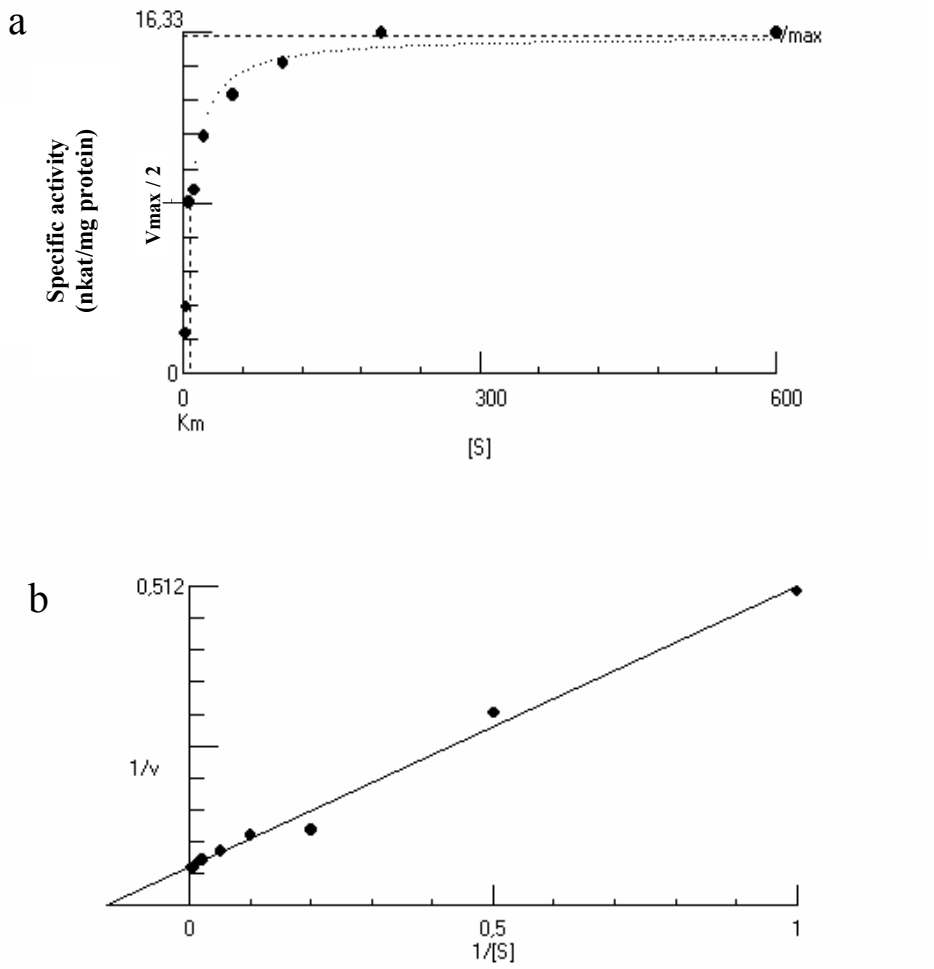


Figure 43. (a) Dependence of Sa4CL1 activity on *p*-coumaric acid concentration, (b) Determination of K_m value for *p*-coumaric acid *via* Lineweaver-Burk plot.

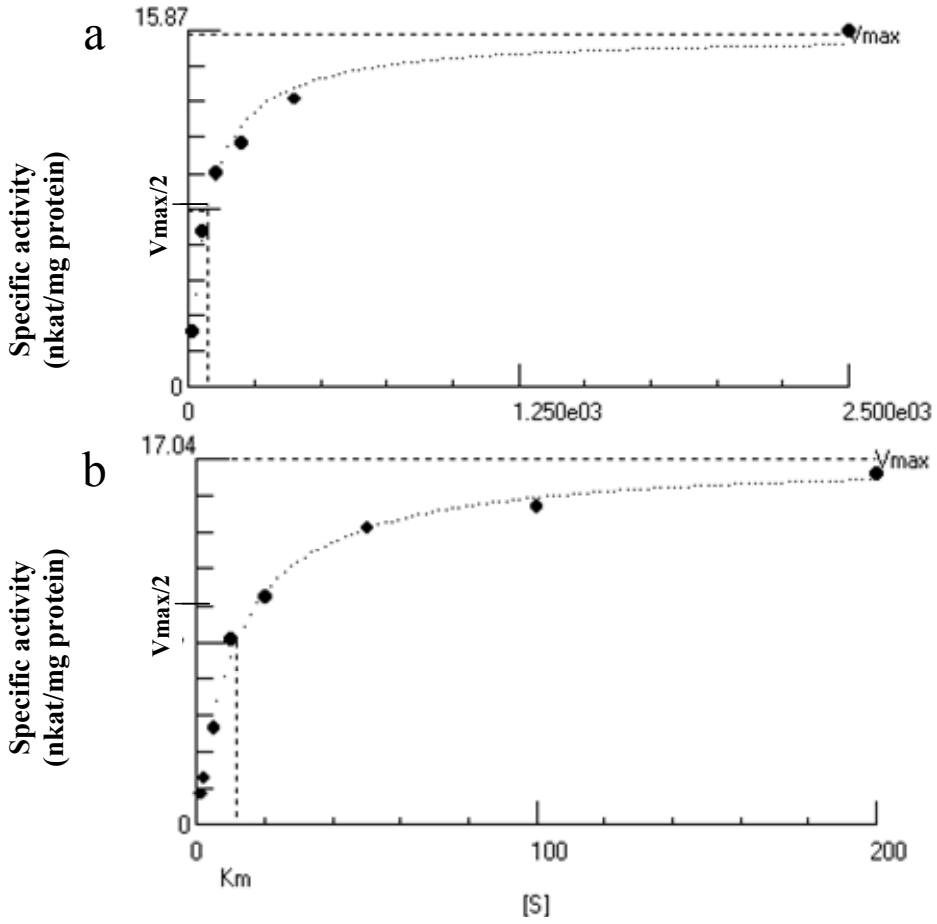


Figure 44. Dependence of Sa4CL1 activity on ATP (a) and CoA (b) concentrations.

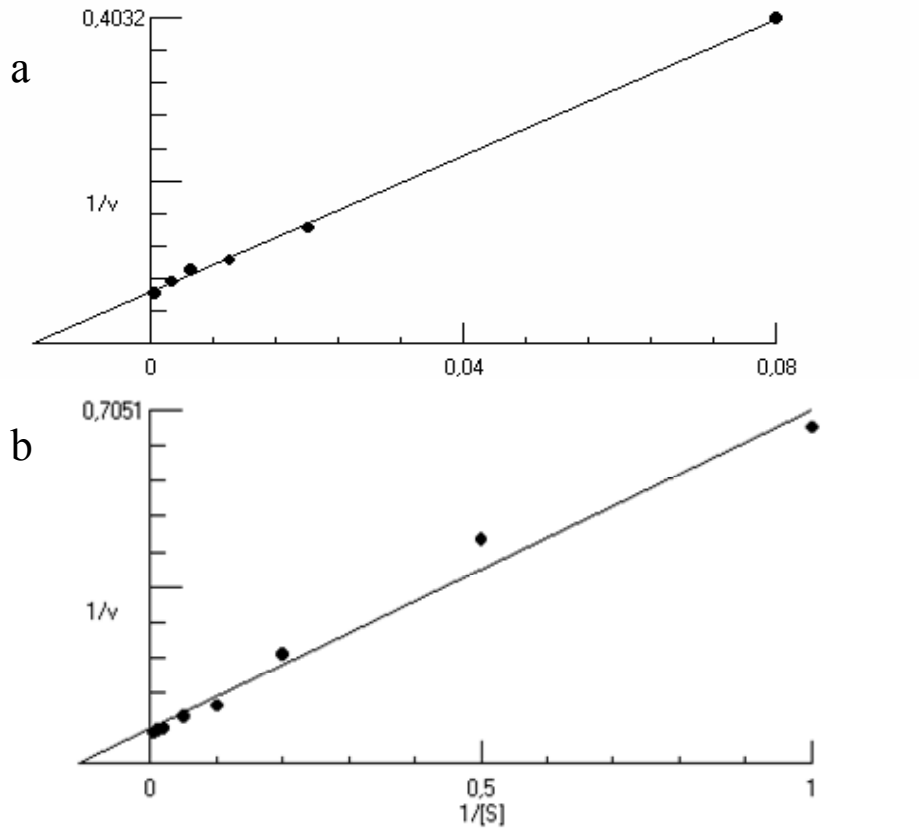


Figure 45. Determination of K_m values for ATP (a) and CoA (b) *via* Lineweaver-Burk plots.

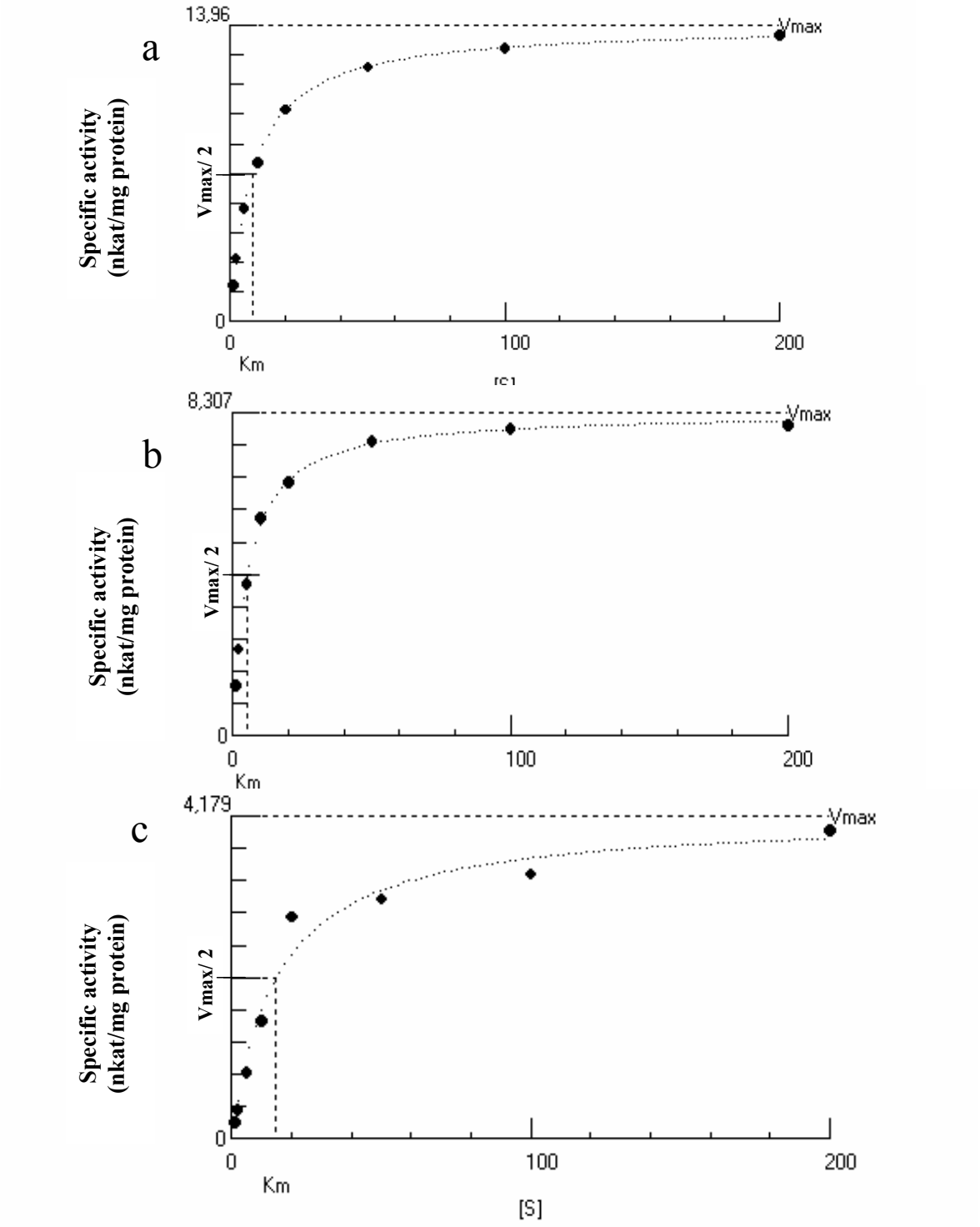


Figure 46. Dependence of Sa4CL1 activity on concentrations of caffeic (a), ferulic (b), and cinnamic acids (c), as illustrated by Michaelis-Menten kinetics.

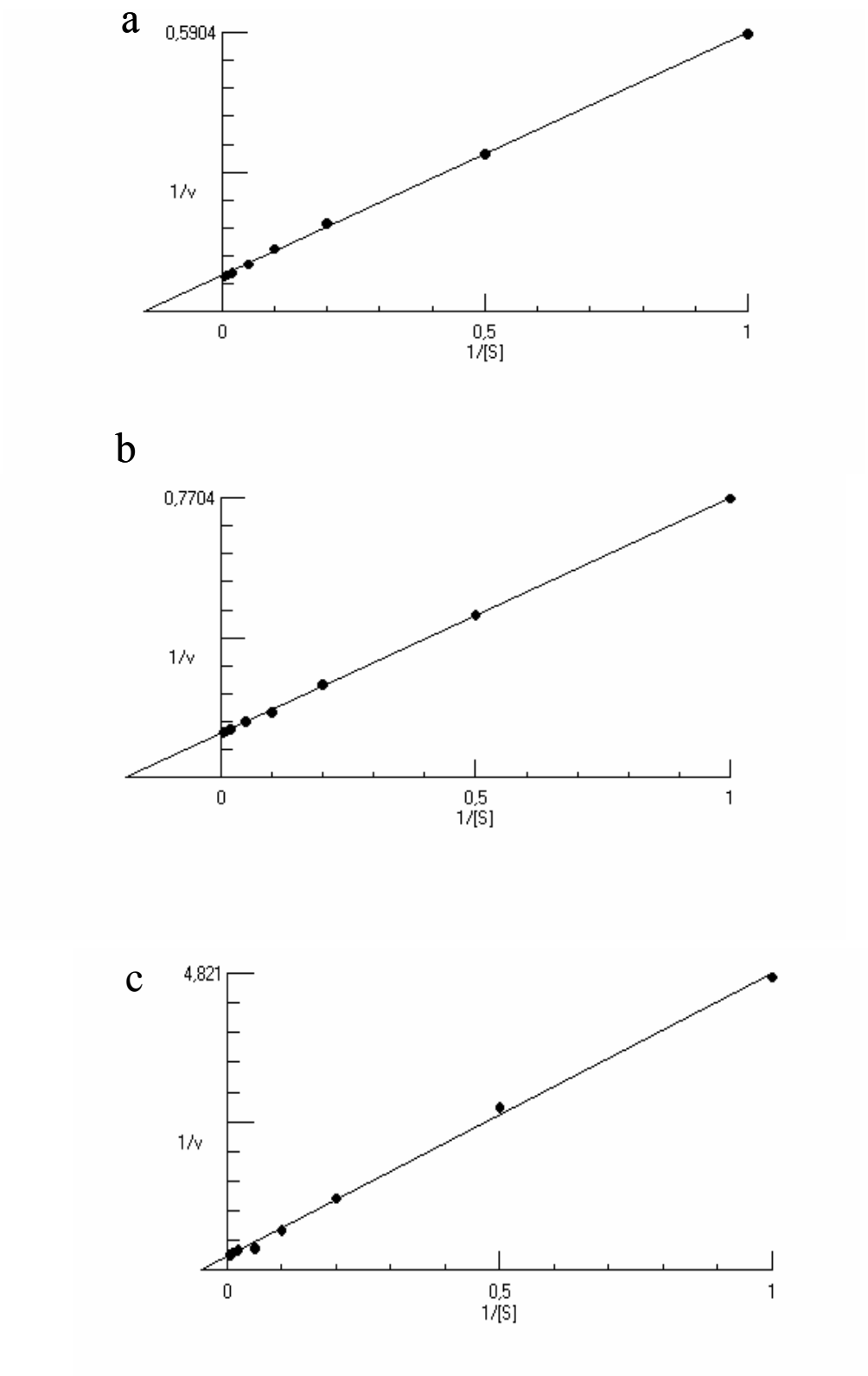


Figure 47. Determination of K_m values for caffeic (a), ferulic (b), and cinnamic acids (c) *via* Lineweaver-Burk plots.

3 Cloning of a cDNA encoding aldehyde dehydrogenase (ALDH)

Sequence data obtained from the new *4CL4* gene fragment (III.2.3.1) were used to derive forward gene-specific primers for attempts to identify the 3´-end of this gene. A 3´-RACE protocol together with forward GSPs (FP1 and FP2, II.1.5.2.2), designed at the coding region of the *4CL4* gene fragment, and adapter primers (RACE long, RACE short; II.1.5.2.5) [Figure 48] amplified the 3´-end of the 4CL1 cDNA only. A 3´end completing the 4CL4 fragment was not obtained, which might be explained by comparatively higher expression of 4CL1 transcripts and high sequence similarity of the designed primers, FP1 and FP2, with the 4CL1 sequence. The 5´-untranslated regions (5'UTR) of the 4CL1 and 4CL4 genes exhibited reasonable differences [Figure 49]. Designing primers in this region may offer a good chance of isolating the 3´-end of the 4CL4 cDNA without 4CL1 cDNA miss-amplification.

Using a forward gene-specific primer, the 5´UTR-primer (II.1.5.2.2), designed in the 5´UTR of 4CL4, was an attempt to get the 4CL4 3´-end. Combination of this primer with an adapter primer, RACE long, and 3´CDS-cDNA (II.3.4.2) as a template led to amplification of a 1.7 kb PCR product. A "Touch down PCR" was used (II.3.5.1) with descending annealing temperature from 55 to 50°C by decreasing 0.5°C per cycle for 10 cycles. Surprisingly, sequence analysis and comparison with sequences available from the databank revealed that the amplified cDNA encoded ALDH. The coding region was 1512 bp including the start and stop codons, ATG and TAA, respectively. The 3´-untranslated region (3´UTR) was about 186 bp. The nucleotide sequence of this SaALDH cDNA is represented in the appendix.

The 1512 bp ORF encoded a 54.8 kDa protein which consisted of 503 amino acids and had an isoelectric point (p*I*) at pH value 5.7. Analysis of the amino acid sequence using the Target P and MITOPROT II programmes (Emanuelsson et al., 2007; Andreoli et al., 2004; Claros and Vincens, 1996) did not identify any N-terminal signal or N-terminal organelle targeting information. The deduced amino acid shows high sequence similarity (79-52%) to ALDHs from castor (accession number B9RKT6), maize (Cui et al., 1996), snapdragon (Long et al., 2009), and tobacco (op den Camp and Kuhlemeier, 1997). An amino acid sequence comparison of various ALDHs is shown in the appendix.

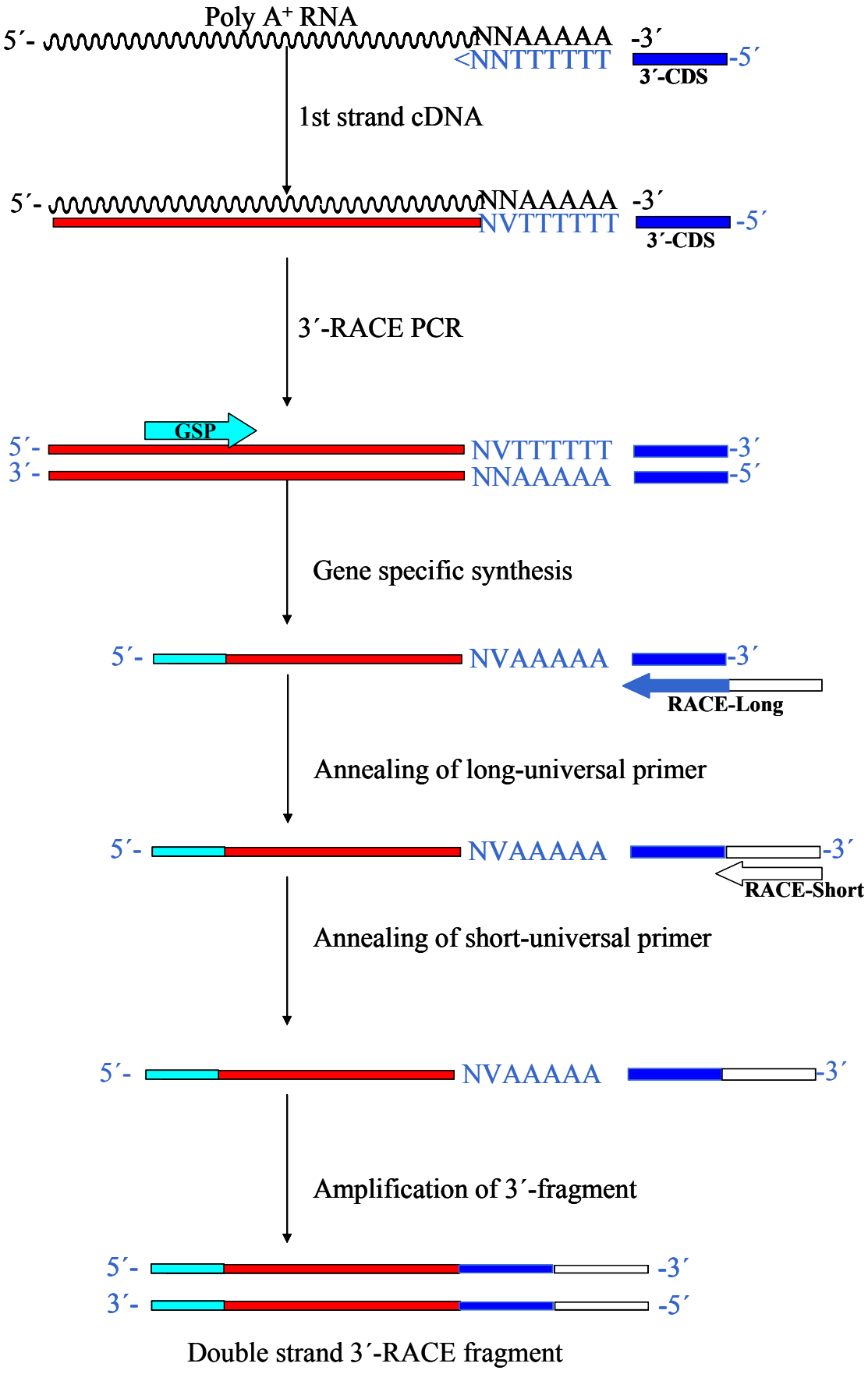


Figure 48. Detailed representation of the 3′-RACE reactions.

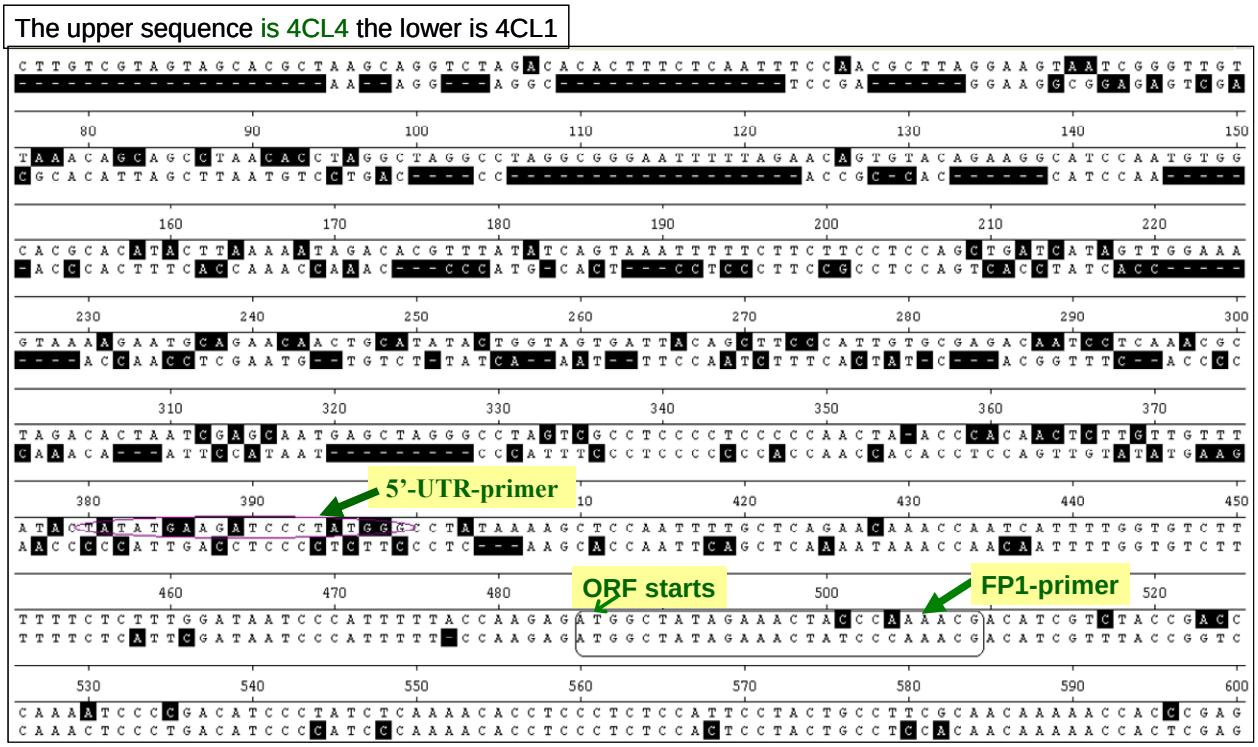


Figure 49. Sequence comparison of 5´-UTRs of 4CL4 (upper) and 4CL1(lower).

4 Cloning of PAL cDNA

For cloning of *S. aucuparia* PAL (SaPAL) cDNA, total RNA was obtained from 9h-elicited cell suspension cultures (II.3.2). Primer pairs (II.1.5.2.8) derived from the databank-available non-coding regions and coding region of the *Malus* and *Pyrus* sequences, respectively, were used to amplify SaPAL cDNA. Annealing temperatures used were 55°C with primer pairs derived from the *Malus* sequence and 60°C in case of primers derived from the *Pyrus* sequence. A PCR product of around 2200 bp was obtained. Sequence analysis revealed a full-length SaPAL cDNA. It contained 2227 nucleotides with an ORF of 2160 nucleotides which encoded a protein of 720 amino acids with a predicted molecular mass of 78 kDa and a p*I* of 6.8. The deduced amino acid sequence showed 97, 96, 92 and 88% sequence similarity to PALs from *Pyrus communis* (DQ230992*), Pyrus x bretschneideri* (GU906268, isoform 1)*, Prunus avium* (O64963, isoform 1), and *Rubus idaeus* (AF237955, isoform 2), respectively. The nucleotide sequence of the SaPAL cDNA is represented in the appendix.

5 Expression analysis by RT-PCR

5.1 Analysis of SaALDH expression

SaALDH mRNA levels were analyzed by RT-PCR. Gene-specific primers (II.1.5.2.7) capable of amplifying a segment of the SaALDH-ORF from 211 to 759 were used. Fixed amounts of total RNA (1 µg) from elicitor-treated and control *S. aucuparia* cell suspension cultures were reverse-transcribed into cDNA (II.3.4) using the 5'-CDS universal primer and the H Minus M-MuLV RT enzyme. The cDNA was diluted 1:10 [Figure 50 a] and 1:20 [Figure 50 b,c] and these samples were used as templates for RT-PCR. A Touch down PCR programme (II.3.5.1) was used and the number of the amplification cycles was optimized. Three different PCR programs, in which amplification cycles were decreased to 25 [Figure 50 a], 20 cycles [Figure 50 b] and 15 cycles [Figure 50 c], were used. The annealing temperature was decreased in 10 cycles from 65-60°C. The use of SaALDH gene-specific primers resulted in amplification of a 548 bp fragment. To control the efficiency of the elicitation process, inducible BIS3 transcript was also determined. The gene specific primers (II.1.5.2.7) capable of amplifying a 1313 bp fragment of the BIS3 cDNA (Liu et al., 2007 & 2010) were the forward primer at the 5´UTR (5´UTR BIS3, also binding to BIS1 cDNA) and the reverse primer at the 3'UTR (3´UTR BIS3). The annealing temperature was decreased from 55-50°C in 10 cycles using a Touch down PCR programme (II.3.5.1).

The masking effect of exponential increase in PCR products through rounds of amplification diminishes the difference in the SaALDH transcription levels before and after elicitor induction [Figure 50 a, b]. Number of amplification cycles and template dilutions were varied, where the optimum conditions were 1:15 and 1:20, respectively. After elicitor treatment, accumulation of BIS3 transcripts reached a maximum level after 9h (Liu et al., 2010). The same induction kinetics was also observed for SaALDH transcripts under optimum cycling condition [Figure 50 c]. The transcription level of SaALDH was higher than that of SaBIS3, which is in agreement with the changes in the specific BD and BIS activities after elicitor treatment (III.1.1).

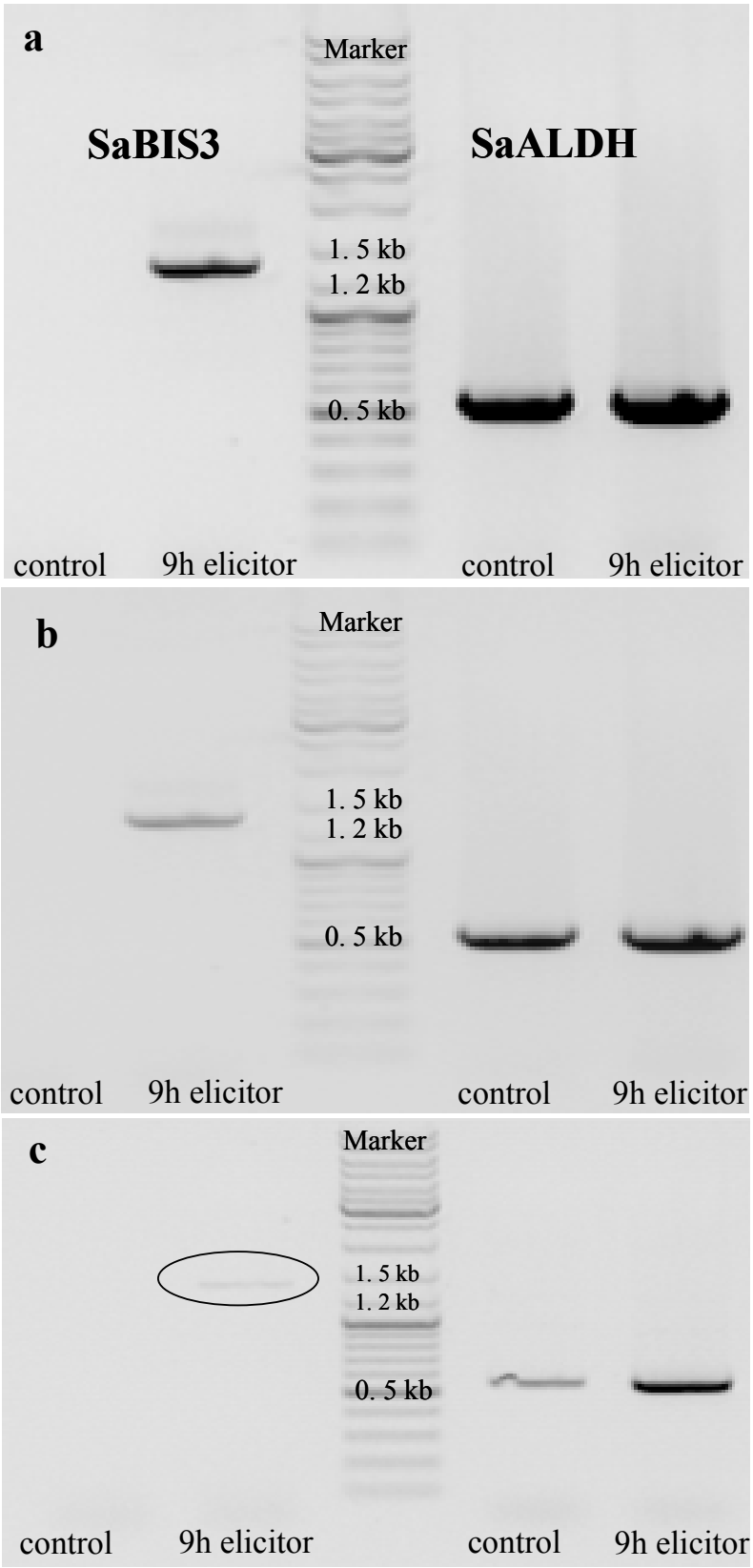


Figure 50. RT-PCR amplification of transcripts encoding SaBIS3 and SaALDH from 9h-chitosan-treated and untreated control cell cultures of *S. aucuparia*. (a) 1:10 template dilution and 25 amplification cycles, (b) 1:20 template dilution and 20 amplification cycles, (c) 1:20 template dilution and 15 amplification cycles.

5.2 Effect of elicitor treatment on expression of Sa4CLs

Transcripts for BIS1 and BIS3, which are elicitor-inducible, start to accumulate 3h after elicitation and reach maximum levels after 6 to 9h (Liu et al., 2007, 2010). Similar accumulation kinetics was observed for PAL transcripts. Therefore, the 9-hour-interval following the onset of elicitation was chosen to study changes in the mRNA levels of the three Sa4CL isoenzymes by RT-PCR. Gene-specific primers (II.1.5.2.7) which selectively amplified fragments of the Sa4CL cDNAs were designed [Figure 51]. The fragment sizes for Sa4CL1, Sa4CL2, and Sa4CL3 were 898, 897, and 896 bp, respectively. The levels of transcripts for PAL (III.4) and BIS1 served as a control for the efficiency and the timing of the elicitation process. The *PAL* and *BIS1* gene-specific primers (II.1.5.2.7) led to amplification of 1070 and 1173, respectively, bp cDNA fragments [Figure 52]. In RNA pools from untreated control cell cultures of *S. aucuparia*, no transcripts for PAL, BIS1, 4CL1, and 4CL2 were detectable, not even when 1 µg template was used for RT-PCR. Low amounts of transcripts were found for Sa4CL3. At various times (3, 6, and 9h) after the addition of chitosan to the cell cultures, total RNA pools were isolated and examined for changes in the transcript levels of interest. Actin transcript contents served as a control for equal RNA template amounts [Figure 52]. PAL and BIS1 transcripts started to accumulate after 3h and reached maximum levels after 6h. Elicitor treatment also induced expression of all three *Sa4CL* genes. The transcript levels reached maxima after 6-9h. However, these maximum mRNA levels of the Sa4CLs were, although somewhat different among each other, significantly lower than the maximum transcript levels observed for PAL and BIS1. When treatment of *S. aucuparia* cell cultures with elicitor was replaced with exposure to light, accumulation of PAL and Sa4CL3 transcripts was observed [Figure 52]. Traces of Sa4CL1 and Sa4CL2 transcripts were only detectable when high amounts of RNA template (1 µg) were used. No light-stimulated expression of the *BIS1* gene was detectable.

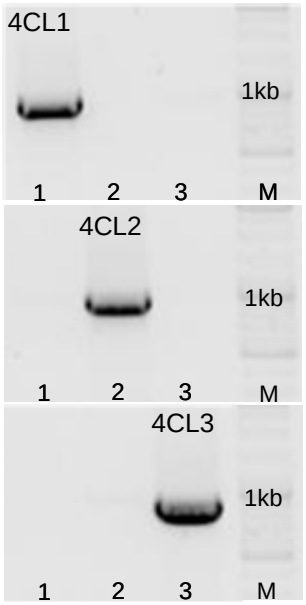


Figure 51. Specificity of the gene-specific primer pairs used to amplify cDNA fragments of Sa4CL1, Sa4CL2, and Sa4CL3. M, marker.

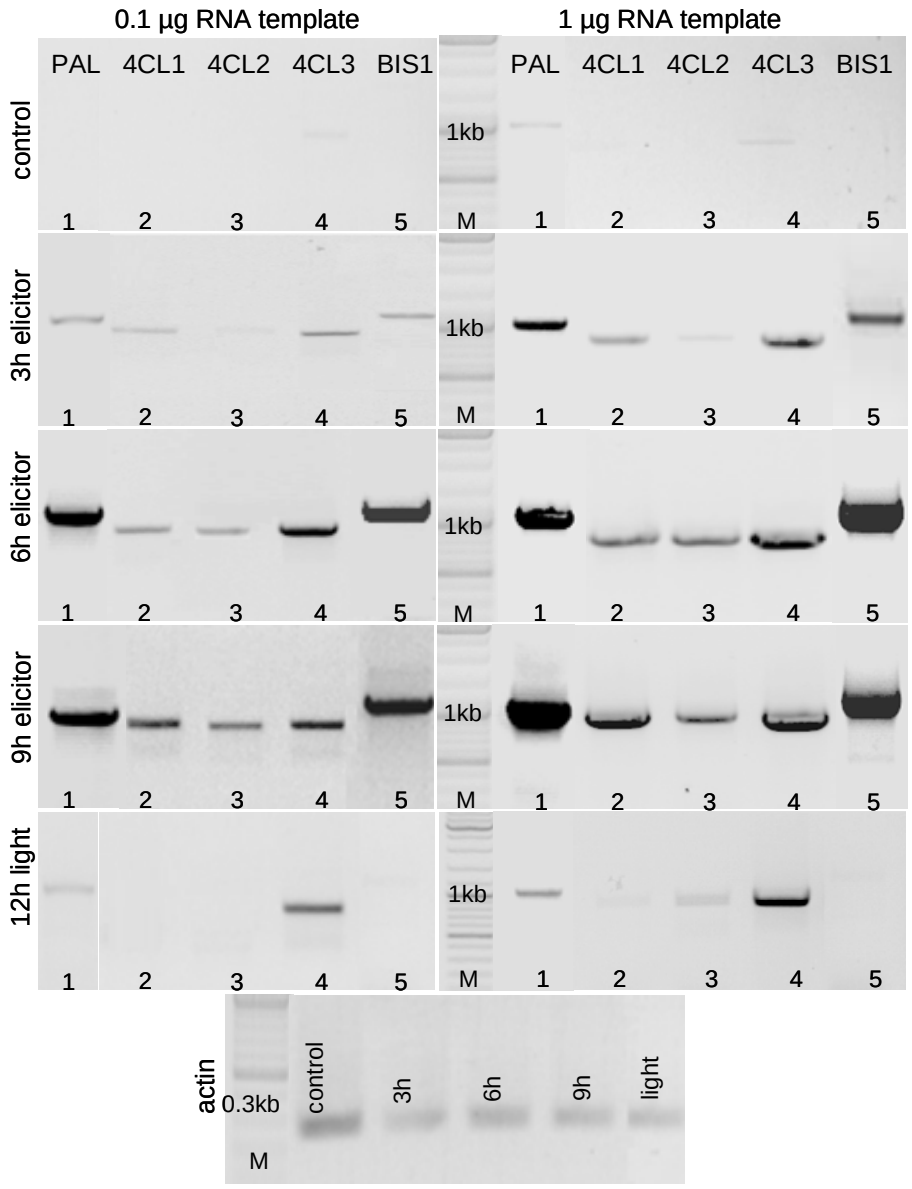


Figure 52. RT-PCR amplification of transcripts encoding PAL (1), Sa4CL1 (2), Sa4CL2 (3), Sa4CL3 (4), and BIS1 (5) at various times after elicitor treatment and following light exposure of *S. aucuparia* cell cultures. Actin transcript levels served as control for equal RNA template amounts. M, marker.

IV. Discussion

1 Aldehyde dehydrogenases (ALDHs)

Endogenous aldehydes are common intermediates of a number of metabolic pathways, including metabolism of amino acids, proteins, lipids and carbohydrates (Schauenstein et al., 1977), whereas the major exogenous source of aldehydes are xenobiotics. Aldehydes can react with cellular nucleophiles because of the electrophilic nature of their carbonyl group (Lindahl, 1992). Excess aldehydes will cause deleterious effects on metabolism. ALDHs have been considered as general detoxifying enzymes which eliminate biogenic and xenobiotic aldehydes in a $NAD(P)^+$-dependent manner (Yoshida et al., 1998). ALDHs play also a role in the degradation of aromatic aldehydes formed as a result of lignin breakdown by a soil bacterium (Crawford et al., 1982) and snails (Large and Connock, 1994). Besides aldehyde detoxification, many ALDH isoenzymes possess multiple additional catalytic and non-catalytic functions such as participating in metabolic pathways, binding proteins, or osmoregulation, to mention a few. The presence of *ALDH* genes in the majority of the species suggests that the ALDH superfamily has an ancient origin. More than 555 distinct *ALDH* genes have been identified throughout all taxa (Sophos and Vasiliou, 2003). Based on protein sequence identity, the ALDH superfamily has been categorized into distinct families (Vasiliou et al., 1999).

The plant ALDHs are represented in 12 distinct families: ALDH2, ALDH3, ALDH5, ALDH6, ALDH7, ALDH10, ALDH11, ALDH12, ALDH18, ALDH19, ALDH21 and ALDH22. Four of them (ALDH11, ALDH19, ALDH21 and ALDH22) are unique to plants and ALDH21 is apparently unique to mosses (Kirch et al., 2004; Wood and Duff, 2009). The *Arabidopsis thaliana* genome has been found to contain 14 unique ALDH sequences encoding members of nine ALDH protein families (Kirch et al., 2004). All of these proteins have a conserved cysteine that aligns with the catalytic cysteine of the ALDH group.

In mammals, at least four different forms of the enzyme are known: class 1 is a tetrameric cytosolic enzyme, class 2 is a tetrameric mitochondrial enzyme, class 3 is involved in the detoxification of aldehydes that form during lipid peroxidation, and class 4 is a microsomal enzyme (Lindahl and Petersen, 1991). Some of the family 3 ALDHs are, in addition, expressed in tumors. ALDH from class 3 is a dimeric cytosolic enzyme.

In human, ALDH is the second enzyme of the major oxidative pathway of alcohol metabolism; the first is alcohol dehydrogenase. Two major liver isoforms of this enzyme, cytosolic and mitochondrial, can be distinguished by their electrophoretic mobilities, kinetic

properties, and subcellular localizations. Most Caucasians have two major isozymes, while approximately 50% of Orientals have only the cytosolic isozyme, missing the mitochondrial isozyme. A remarkably higher frequency of acute alcohol intoxication among Orientals than among Caucasians could be related to the absence of the mitochondrial isozyme. The mitochondrial isoform has a low K_m for acetaldehydes.

Compared to the comprehensive study of ALDHs in humans (Yoshida et al., 1998; Marchitti et al., 2008), only a small number of plant ALDHs have been functionally characterized (Kirch et al., 2004). To date, most of the studied plant *ALDH* genes are shown to be induced under high salinity or water deficit conditions, suggesting possible roles of these genes in improving the plant osmotic stress tolerance (Kotchoni and Bartels, 2003; Kirch et al., 2004; Kirch et al., 2005). Several studies indicate that over-expression of some plant ALDHs indeed enhances plant tolerance to diverse abiotic stresses (Kotchoni and Bartels, 2003; Kotchoni et al., 2006; Huang et al., 2008).

The first identified plant *ALDH2* gene rf2 is required for male fertility in maize (Liu et al., 2001). The study of rice ALDH2a shows that this enzyme might be responsible for efficient detoxification of acetaldehyde during re-aeration after submergence of rice plants (Tsuji et al., 2003).

A partial explanation for the presence of so many *ALDH* genes in plant genomes is the need to provide ALDH activity in various subcellular compartments. Although some aldehydes (e.g. acetaldehyde) are able to move from one subcellular compartment to another, the molecular sizes of others preclude their passive diffusion across membranes. Therefore, organelles that contain aldehyde-generating pathways but do not contain an ALDH could experience aldehyde-induced damage.

The enzyme has three domains, a $NAD(P)^+$ binding domain, a catalytic domain, and an “arm-like” bridging domain (Liu et al., 1997). The active site of ALDH is largely conserved throughout the different classes of the enzyme and, although the number of amino acids present in a subunit can change, the overall function of the site changes little. The active site will contain one molecule of an aldehyde and a $NAD(P)^+$ that functions as a cofactor. A cysteine and a glutamate will interact with the aldehyde substrate. Many other residues will interact with the $NAD(P)^+$ to hold it in place.

In $NAD(P)^+$-dependent reaction, the aldehyde enters the active site through a channel located on the outside of the enzyme. The catalytic mechanism [Figure 53] starts with nucleophilic attack of sulfur from a cysteine in the active site on the carbonyl carbon of the aldehyde substrate to form an oxyanion thiohemiacetal intermediate. The hydrogen is kicked off as a

hydride and attacks NAD(P)$^+$ to make NAD(P)H. The enzyme's active site then goes through an isomorphic change where the NAD(P)H molecule is moved, creating room for a water molecule to access the substrate. The water is primed by a glutamate in the active site, and makes a nucleophilic attack on the carbonyl carbon, kicking off the sulfur as a leaving group. The active site contains a Rossmann fold and interactions between the cofactor and the fold allow for the isomerization of the enzyme while keeping the active site functional (Liu et al., 1997).

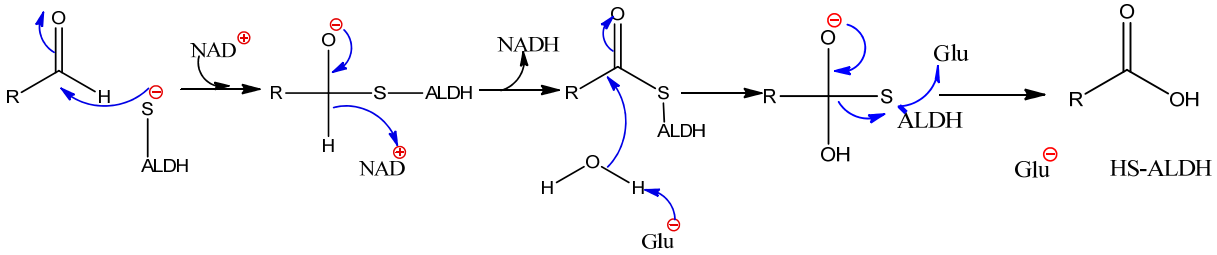


Figure 53. General reaction mechanism of ALDHs.

1.1 *S. aucuparia* ALDH (SaALDH)

The SaALDH ORF encodes a 54.8 kDa protein as a mature subunit that consists of 504 amino acids. The active ALDH enzyme is generally either homodimeric or homotetrameric (Rodriguez-Zavala and Weiner, 2002). Gel-filtration chromatography has to be done to determine the SaALDH native molecular mass and whether the enzyme is homodimer or homotetramer. A tetrameric structure is common for members of the ALDH2 family in mammals and yeast (Yoshida et al., 1998). In plants, only two maize mitochondrial enzymes (RF2A and RF2B) and one benzaldehyde dehydrogenase from snapdragon, out of the limited number of characterized plant ALDHs (Liu and Schnable, 2002; Long et al., 2009), exist as native tetrameric proteins. Analysis of the amino-acid sequence of SaALDH did not identify a N-terminal signal or N-terminal organelle targeting information.

Generally, sequence comparisons among *ALDH* genes from bacteria, animals and plants have identified three diagnostic amino acid motifs: (i) the ALDH glutamic acid active site signature; (ii) the Rossmann fold GxGxxG coenzyme binding site; and (iii) the catalytic thiol (Vasiliou et al., 1999; Perozich et al., 1999). All the conserved amino acids characteristic for ALDHs are present in SaALDH [Figure 54].

The active site signature Leu-**Glu**-Leu-Gly-Gly-Lys-Ser-Pro (L**E**LGGKSP; corresponding to the SaALDH amino acids 269–276) includes the glutamate residue (E^{270}, Figure 53) as the important amino acid (Wang and Weiner, 1995). The cysteine at position 304 (C^{304}) serves as the catalytic residue (Farres et al., 1995) and the histidine at position 237 (H^{237}) is necessary for the correct folding of the native ALDH (Zheng and Weiner, 1993). A sequence comparison of SaALDH with other ALDHs is shown in the appendix.

Ser^{74} (Rout and Weiner, 1994) and Glu^{487} (Farres et al., 1994) present in rat liver mitochondrial ALDH, both with a proposed function in NAD^+ binding, are not conserved in the ALDHs from tobacco, maize and snapdragon. However, Ser^{74} is conserved in *S. aucuparia* (S^{34}) and castor (S^{32}) ALDHs, while Glu^{487} is conserved in neither gene. It was concluded that the role of residue Glu^{487} in human liver mitochondrial ALDH could be to provide hydrogen or salt bonding to the coenzyme (Farres et al., 1994).

MGSDFNGTSASDSFVKTPTIKFTQLFINGEFLDSVSGKTFETIDPRTGDVVTRV
AEGDKEDVDLAVKAARAAFDHGSWPRLPGAERGRIMMKFADLIDNMNEELAIL
DTVDAGKLFSMGKTIDIPQVAEMVRYYAGAADKIHGEVLKMSRELHGYTLLEPI
GVVGLIVPWNFPSTLLFGKVSPALAAGCTMVIKPAEQTPLSALYYAHLAKLAGV
PDGVLNVITGFGKTAGAAISS**H**MDIDKVSFTGSTEVGRLVMQAAAKSNLKPVS
LELGGKSPLVIFDDADINMAADLALLGILYNKGEI**C**VASSRVYVQEGIYDEFVKK
LQEKAKDWVVGDPFDPNVRQGPQVDKKQFEKILSYIEHGKKEGATLLTGGKP
VGNKGYYIEPTIFTDVKDDMVIAQDEIFGPVLALMKFKTIEEAIQRANNTRYGLA
AGIITKDLNVANTVSRSIRAGIIWINCYFAFDRDCPYGGYKMSGFGRDFGMQGL
YHYLHTKSVVTPLFNSPWL

Figure 54. Full-length amino acid sequence of SaALDH with conserved residues characteristic for ALDHs.

It was previously shown that ALDH activity is necessary for tobacco pollen growth (op den Camp and Kuhlemeier, 1997). Acetaldehyde is one of the intermediate products of ethanolic fermentation, which is highly active in germinating pollen. It will be interesting to find out if SaALDH is involved in detoxification of acetaldehyde. First, the inhibitory effect of disulfiram, an inhibitor of ALDHs (Kitson, 1975), on pollen germination should be examined. Then, RT-PCR can be used to detect SaALDH expression in *S. aucuparia* pollen grains before and after germination, suggesting that pollen germination requires the function of ALDH. Absence of SaALDH transcripts before and after germination will rule out involvement of SaALDH in acetaldehyde detoxification and open a question for its role in benzoic acid biosynthesis.

1.2 SaALDH expression

In contrast to known plant ALDHs, which are expressed constitutively (Tsuji et al., 2003), SaALDH transcripts were only found at a basal expression level in untreated cell cultures, but were up-regulated upon chitosan elicitation. The up-regulation of SaALDH expression after induction suggests its involvement in benzoic acid biosynthesis because all benzoic acid biosynthetic enzymes should be coordinately induced to form the biphenyl defense compound aucuparin as a major component. After induction, the expression level of SaALDH transcripts is much higher than that of SaBIS3 transcripts, which is also true at the biochemical level for BIS and BD. Previously, a similar difference between BD and BPS, the key enzyme of xanthone biosynthesis, was observed in cell cultures of *H. androsaemum* (Abd El-Mawla and Beerhues, 2002). At the biochemical level, maximum BD activity detected in crude protein extracts from *S. aucuparia* (17h post-elicitation) was 3.5 times that of control cells (Gaid et al., 2009). The same difference was observed with SaALDH mRNA before and after induction. Inhibition of phytoalexin production as well as enhancement of benzaldehyde accumulation upon silencing of the *SaALDH* gene (Ossowski et al., 2008) together with the kinetic analysis of the recombinant protein and expression studies will give more information about the role of SaALDH in benzoic acid biosynthesis.

2 Benzaldehyde dehydrogenase (BD)

Benzoic acid biosynthesis in plants remains an important unresolved question despite numerous attempts at its elucidation. Several side chain-shortening routes have been proposed for BA formation from L-Phe; however, no genes or enzymes responsible for these biochemical steps have been discovered (Wildermuth, 2006).

Enzymes that can catalyze the conversion of benzaldehyde to benzoic acid have been reported from bacteria and animals (Yoshida et al., 1998; Yasuhara et al., 2002; Hirano et al., 2007). In plants, BD was isolated and biochemically characterized from *Antirrhinum majus* (snapdragon), whose flowers have been shown to emit high levels of methylbenzoate, the product of benzoic acid carboxyl methylation. The protein is a mitochondrial homo-tetrameric enzyme that belongs to the plant ALDH2 family (Long et al., 2009). In contrast to known plant ALDHs, which are expressed constitutively (Tsuji at al., 2003) or in a wide variety of organs (Liu and Schnable, 2002), BD mRNA in *A. majus* was found exclusively in the scent parts of flowers (Long et al., 2009). Another recent report was about aldehyde oxidase (AAO4) from *A. thaliana* siliques. The partially purified protein showed broad substrate specificity. Among the substrates tested, the lowest K_m value (23 μM) was observed with

benzaldehyde (Ibdah et al., 2009). ALDH2C4 in *Arabidopsis* is involved in ferulic and sinapic acids biosynthesis (Nair et al., 2004). Maize recombinant aldehyde oxidases RF2a and RF2b both can oxidize a series of aliphatic saturated aldehydes. However, benzaldehyde can only be oxidized with RF2a (K_m = 3.4 µM; Liu and Schnable, 2002). Aldehyde oxidase (AAO3) from *A. thaliana* prefers abscisic aldehyde (K_m = 0.51 µM) and was found to also oxidize benzaldehyde (K_m = 44 µM) (Seo et al., 2000). Recombinant *A. thaliana* aldehyde oxidase (AOα) showed high affinity for benzaldehyde and indole-3-aldehyde (Koiwai et al., 2000). *Hordeum vulgare* (barley) roots contains an aldehyde oxidase (AO4), which exhibited high affinity toward heptaldehyde, acetaldehyde and benzaldehyde (Omarov et al., 1999). Purified maize AO exhibited relatively broad substrate specificity, good substrates among the tested compounds were indole-3-aldehyde and benzaldehyde (Koshiba et al., 1996). Among all reports for ALDHs, only one cytoplasmic AAO4 whose main function is to convert benzaldehyde to BA has been reported (Ibdah et al., 2009).

2.1 BD from *S. aucuparia* cell cultures

The starter substrate for BIS (I.3.3) is benzoyl-CoA. The formation of benzoyl-CoA proceeds *via* benzaldehyde as an intermediate. NAD^+ -dependent benzaldehyde dehydrogenase (BD; EC 1.2.1.28) catalyzes the last reaction of benzoic acid biosynthesis and converts benzaldehyde to benzoic acid [Figure 55].

In *S. aucuparia* cell cultures, BIS expression was rapidly, strongly and transiently induced by elicitor treatment resulting in a massive accumulation of aucuparin, the best-known biphenyl phytoalexin (Liu et al., 2004). Cultured cells of *Lithospermum erythrorhizon* and elicitor-treated cell cultures as well as hairy root cultures of *Daucus carota* (Yazaki et al., 1991; Schnitzler et al., 1992; Sircar and Mitra, 2008) accumulate 4-hydroxybenzoic acid. Accordingly, the preferred substrate for the *L. erythrorhizon* and *D. carota* dehydrogenases was 4-hydroxybenzaldehyde. Benzaldehyde was only a poor substrate for these enzymes. Thus, *S. aucuparia* cell cultures appear to be a valuable system for studying the biosynthesis of benzoic acid *via* benzaldehyde.

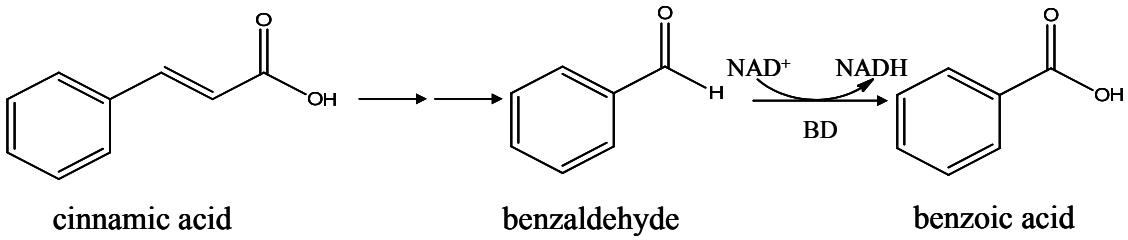


Figure 55. Benzaldehyde dehydrogenase (BD)-catalyzed formation of benzoic acid from benzaldehyde, itself derived from cinnamic acid.

Detection of BD activity in *S. aucuparia* cell cultures was suggestive of a benzoic acid biosynthetic route *via* benzaldehyde. The properties of BD were studied in desalted cell-free extracts. Detection and characterization of this enzyme provides first insight into benzoic acid metabolism in the economically valuable taxon Pyrinae (formerly Maloideae). BD and BIS activities were coordinately induced upon chitosan addition, indicating that the BD-catalyzed reaction is involved in the biosynthesis of aucuparin. These results agreed with recently published RNA blot analysis for BIS transcripts (Liu et al., 2007). Similar to other plant ALDHs, whose substrate specificity is relatively broad, the basal level of dehydrogenase activity at early time points after the onset of elicitation might be due to ALDHs from primary metabolism that can poorly accept benzaldehyde. However, the difference between the specific BD activities before and after induction is similar to the difference observed at transcription level for SaALDH, suggesting that BD activity at early time points is related to the same enzyme activity detected at later post-elicitation times. The same co-ordinate induction of BD and BPS was observed in methyl-jasmonate-treated cell cultures of *H. androsaemum* (Abd El-Mawla and Beerhues, 2002). Like biphenyls, xanthones are formed from benzoyl-CoA and three molecules of malonyl-CoA. BIS and BPS are characterized by similar substrate but different product specificities (Liu et al., 2007). The substrate specificity of *H. androsaemum* BD resembled that of the enzyme from *S. aucuparia,* although the studies of *H. androsaemum* BD included only benzaldehyde and monohydroxybenzaldehydes (Abd El-Mawla and Beerhues, 2002). Furthermore, the specific activity of *H. androsaemum* BD was markedly higher than that of BPS, which is similar to the ratio of the specific activities of *S. aucuparia* BD and BIS.

BD activity from *S. aucuparia* cell cultures was not stimulated by the presence of divalent metal cations, which resembles the observations with previously studied BD activity from *D. carota* hairy roots (Sircar, 2009) and benzaldehyde oxidase activity from *A. thaliana* siliques (Ibdah et al., 2009). The apparent K_m values for the BD substrates from *S. aucuparia* cell

cultures were in the range of previously reported values (Schnitzler et al., 1992; Sircar, 2009). The preferred substrate for BD was benzaldehyde. Salicylaldehyde and other substituted aldehydes were poor substrates. Vanillin and trihydroxybenzaldehyde failed to give any detectable activity. These results are in agreement with recent published data for BD from *A. majus* (snapdragon; Long et al., 2009). Recently, accumulation of 4-hydroxycoumarin in the culture medium of *S. aucuparia* suspensions was detected upon feeding the elicitor-treated cell cultures with the *N*-acetylcysteamine thioester of salicylic acid (salicoyl-NAC) (Liu et al., 2010), the physiological ester of salicylic acid. The CoA thioester was unlikely to be taken up into the cultured cells due to its size and charge. Although BIS1 (Liu et al., 2007) and BIS3 (Liu et al., 2010), which are the elicitor-induced isoenzymes, accept and prefer, respectively, salicoyl-CoA in *in vitro* assays [Figure 56], the accumulated phytoalexin in elicitor-treated *S. aucuparia* cell cultures is the benzoyl-derived biphenyl aucuparin. 4-Hydroxycoumarin was not detectable in these cell cultures, indicating that salicoyl-CoA is not available as an endogenous starter substrate for the BIS isoenzymes. In the present work, *S. aucuparia* BD exhibited only low relative activity with salicylaldehyde, likewise pointing to a minor role of the 2-hydroxylated intermediates in benzoic acid metabolism of elicitor-treated *S. aucuparia* cell cultures.

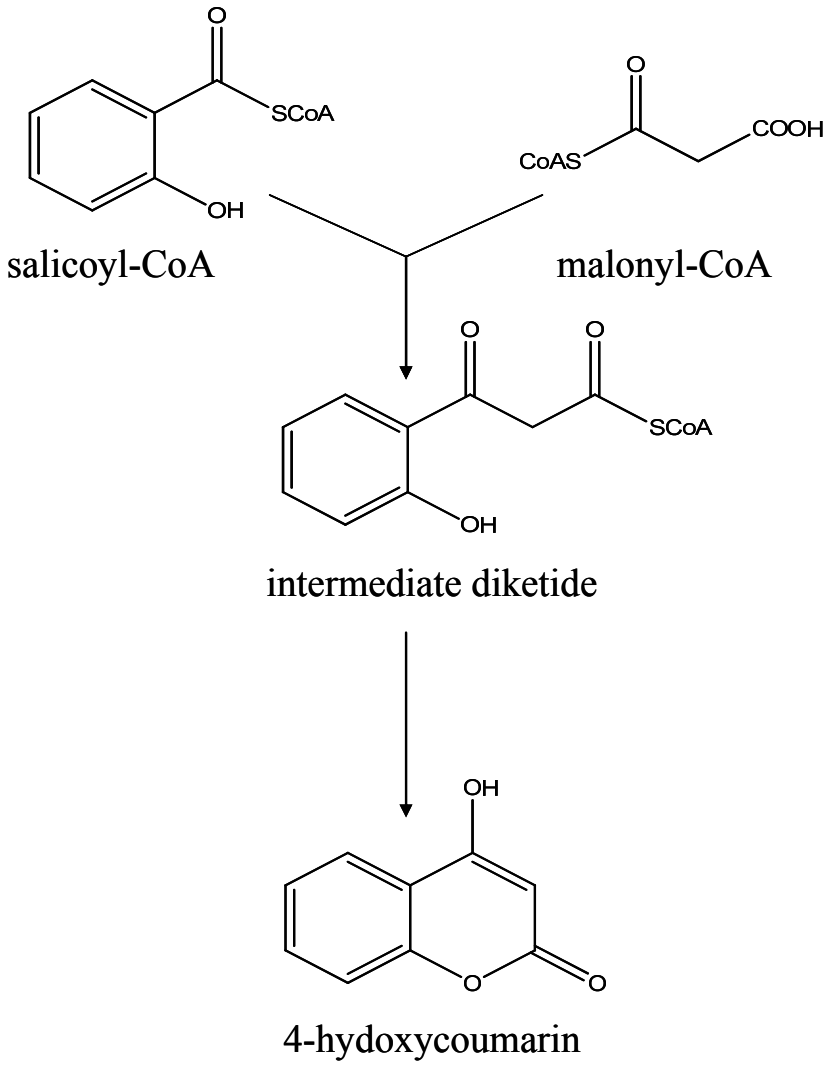


Figure 56. BIS-catalyzed formation of 4-hydroxycoumarin from salicoyl-CoA and malonyl-CoA (Liu et al., 2010).

Establishing the *BD* gene function at molecular level is important for manipulation of benzoate- primed biosynthetic pathways (e.g. biphenyl biosynthesis in Pyrinae). Phytoalexin-producing plants, such as *S. aucuparia*, *Eriobotrya japonica*, and *Photinia glabra* whose defence compounds, biphenyls and dibenzofurans, accumulated in leaves (Miyakodo et al., 1985; Widyastuti et al., 1992; Kokubun and Harborne, 1994, 1995; Hrazdina, 2003), can be used to assess the biological function of *S. aucuparia* BD *in vivo*. Metabolically engineered plants expressing BD might produce a higher level of BA, which is a precursor for pathogen-induced phytoalexins resulting in a more resistant plant line.

3 CoA thioesters

CoA thioesters represent an important class of activated intermediates in various biological pathways. This type of activation can facilitate the transfer of the acyl moiety by enzymes known as acyltransferases (Mishra and Drueckhammer, 2000; St-Pierre and De Luca, 2000). Additionally, acylated intermediates can participate in reductive reactions catalyzed by oxidoreductases (Lacombe et al., 1997) and in aldol-type reactions catalyzed by Claisen enzymes (Mishra and Drueckhammer, 2000). Unlike prosthetic groups, which function by remaining bound to a single enzyme, CoA acts as a diffusible carrier of acyl groups from one enzyme-catalyzed reaction to another. While the structure of CoA is fairly complex, it is functionally a simple molecule. The enzymatic reactions of CoA involve only the thiol group, while reactions of CoA thioesters involve the thioester group and/or the acyl moiety. The rest of the CoA molecule serves as a recognition element for binding by enzymes.

4 *p*-Coumarate-CoA ligases

The acyl-activating capacity of CoA ligases, which belong to the superfamily of adenylate-forming enzymes, is believed to constitute a central function in the biosynthesis of a host of both primary and secondary compounds in various pathways. Among these, the biosynthesis of many phenylpropanoid compounds in green land plants involves the activation of *p*-coumarate and a number of structurally related substrates [Figure 57] to the corresponding CoA thioesters, which is catalyzed by *p*-coumarate-CoA ligase (4CL; EC 6.2.1.12; Cukovic et al., 2001). Channeling of photosynthetically fixed carbon through the phenylpropanoid pathway in plants requires the sequential action of three enzymes: phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4CL, which together constitute the general phenylpropanoid pathway.

Because of the importance of phenylpropanoid-derived compounds in plants, 4CL has been the subject of extensive study for many years, mainly in higher plants. It has been shown to occur in the form of multiple isoenzymes with either similar substrate affinities like in tobacco (*Nicotiana tabacum*; Nt4CL; Lozoya et al., 1988) or with different substrate affinities and/or tissue distribution like in soybean (*Glycine max*; Gm4CL; Lindermayr et al., 2002).

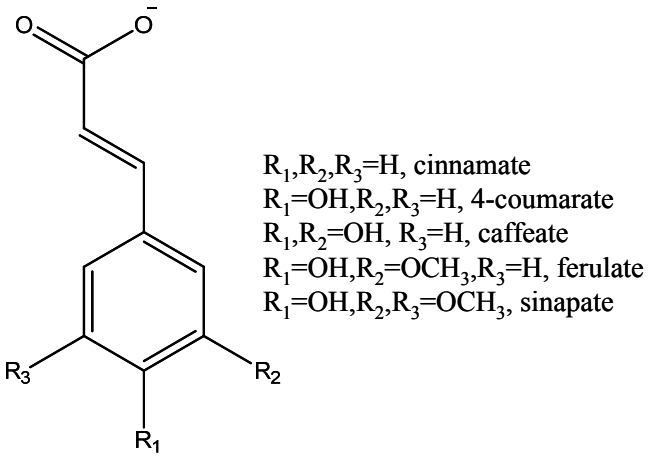


Figure 57. Chemical structures of the five naturally occurring 4CL substrates.

It has been suggested that one role for multiple forms of 4CL might be channeling carbon through various sub-branches within phenylpropanoid metabolism through specific interactions amongst different isoforms of the three enzymes that make up the core phenylpropanoid pathway. Direct biochemical evidence for such interaction is limited, and most efforts have been focused on PAL and C4H rather than 4CL (Rasmussen and Dixon, 1999; Achnine et al., 2004).

Based on the structural relationships between 4CLs from different plants, the existence of two functionally divergent classes of enzyme has been proposed (Ehlting et al., 1999). The *Arabidopsis* class I isoforms, At4CL1 and At4CL2, are constitutively expressed in lignified stems of adult plants and at the onset of lignin deposition in seedlings, roots and cotyledons (Ehlting et al., 1999; Lee et al., 1995). Class I 4CL isoforms from aspen (*Populus tremuloides;* Pt4CL1) and tobacco are likewise highly expressed in lignifying tissues (Hu et al., 1998; Lee and Douglas, 1996). Contrasting expression patterns have been found for class II 4CL isoforms. For example, At4CL3 from *A. thaliana* is expressed in all light-exposed organs such as leaves, flowers, and siliques, and Pt4CL2 from aspen is restricted to epidermal cells of leaves and stem, suggesting that both enzymes are associated with nonlignin-related phenylpropanoids that function as UV protectants (Ehlting et al., 1999; Hu et al., 1998). Sequence analysis, phylogenetic comparisons and expression study of *Sorbus* 4CL isoforms suggest that Sa4CL1 and Sa4CL2 (accession numbers GU938595 and GU949552, respectively) belong to class I, whereas Sa4CL3 (accession numbers GU949553) corresponds to class II (Ramadan, 2006; Scharnhop, 2008; this work).

The presence of a highly conserved putative AMP-binding domain has been used as the most important criterion to group enzymes such as 4CLs, firefly luciferases, acetyl-CoA synthetases, fatty acyl-CoA synthetases and nonribosomal peptide synthetases in one superfamily of adenylate-forming enzymes. Despite low overall amino acid sequence identity shared by these enzymes, similar reaction mechanisms and the presence of conserved peptide motifs were the main criteria to group these enzymes together into one superfamily (Fulda et al., 1994). Within the superfamily of adenylate-forming enzymes, phylogenetic analyses showed that 4CL forms a monophyletic plant-specific group that is most closely related to luciferases, when compared with long chain acyl-CoA synthetases and acetyl-CoA synthetases (Cukovic et al., 2001).

For three members of the above-described superfamily of adenylate-forming enzymes, the crystal structures have been elucidated. Firefly luciferase crystals were obtained in the absence of its substrate luciferin (Conti et al., 1996). Of the nonribosomal peptide synthetases, the phenylalanine-activating domain (PheA) of gramicidin S synthetase from *Bacillus brevis* (Conti et al., 1997) and the 2,3-dihydroxybenzoic acid-activating enzyme from *Bacillus subtillis* (May et al., 2002) were cocrystallized with ATP and their specific substrates phenylalanine and 2,3-dihydroxybenzoate, respectively. The crystal structure of acetyl:CoA synthetase from *Salmonella enterica* in complex with both CoA and adenosine-5'-propylphosphate, an inhibitor of the related propionyl-CoA synthetase, has also been reported (Gulick et al., 2003).

The structures of both PheA and luciferase are characterized by a small C-terminal domain important for catalytic activity (adenylation) and a larger N-terminal domain with a structure based on two similar *β*-sheets and one *β*-barrel. Amino acids within the corresponding *A. thaliana* 4CL2 (At4CL2) C-terminal region that are conserved with PheA and other adenylate-forming enzymes are also catalytically important (Stuible et al., 2000). Assuming co-linear structures of 4CL and PheA, the two sequences can be superimposed by anchoring the conserved box II (GEICIRG) domain of 4CL to the C3 strand within the β-barrel of the PheA structure (GELCIGG) (Ehlting et al., 2001). Within the adenylation domains of bacterial peptide synthetases, amino acid residues that directly bind the amino acid substrate and confer substrate specificity for amino acid adenylation were identified using information from the PheA crystal structure, primary and secondary structure comparisons, and site directed mutagenesis (Conti et al., 1997; Stachelhaus et al., 1999).

Protein domains typical of other plant 4CLs (Stuible et al., 2000) are also present in *S. aucuparia* 4CLs [Figure 58]. The AMP-binding motif is highly conserved among plant 4CLs

and other adenylate-forming enzymes (Ehlting et al., 2001). Near the N-terminus is the conserved motif I (LP Y/F SSGTTGPKG), and toward the C-terminus is the second conserved Box II sequence, GEICIRG. Interestingly, the seven amino acids of motif II are fully conserved in two *Rubus idaeus* 4CLs, Ri4CL1 and Ri4CL2. In Ri4CL3, this motif has a slightly variant amino acid sequence, GEICVRG (Kumar and Ellis, 2003). Such a variation was also reported for Lp4CL3 from perennial ryegrass (*Lolium perenne*; Heath et al., 2002).

The Box II motif, GEICIRG, is absolutely conserved in all 4CLs, and its central cysteine residue was suggested to be directly involved in catalysis (Becker-André et al., 1991). This postulation was supported by using a sulfhydryl-modifying agent, which can abolish 4CL activity (Knobloch and Hahlbrock, 1977). However, evidence in conflict with the participation of a cysteine residue in catalysis accumulated for other adenylate-forming enzymes, resulting in a revised model of adenylate formation by peptide synthetases, which excluded the direct involvement of a cysteine residue in this reaction step (Stein et al., 1996). Later, Stuible et al. (2000) disproved a direct involvement of the conserved GEICIRG of 4CLs in their catalytic activity. Two mutations introduced into the Box II motif of *A. thaliana* 4CL2 (Glu401Gln, Cys403Ala) reduced the specific activity to 21 and 45% of the wild-type level, respectively. The point mutant Cys403Ala still exhibited considerable enzymatic activity. These results can exclude the direct involvement of Cys^{403} in adenylate or thiol ester formation. A reduction in activity may indicate a supportive role of this amino acid in catalysis or stabilization of the protein structure.

Met A I E T I P N D I V Y R S K L P D I P I P K H L P L H S Y C L H N K N H S S S K
P C I I D G A T G D I Y T F A D V E L N A R R V A S G L N K L G I Q Q G D V I Met
L L L P N S P A F A F A F A F L G A S F R G A Met T T A A N P F F T P A E I L K Q A K
A S K A K L I I T L A C Y Y D K V K D L S S S S S D D V H D I K L Met C V D S P P
D P S C L H F S E L L Q A D E N D Met P E V D I S P D D V V A L **P Y S S G T T G**
L P K G V Met L T H K G L V T S V A Q Q V D G E N P N L Y Y S T D D V V L C V
L P L F H I Y S L N S V L L C G L R A G A A I L Met Met N K F E I V S L L G L I D
K Y K V S I A P I V P P I V L A I A K F P D L D K Y D L S S I R V L K C G G A P L
G K E L E D T V R A K F P N V T L **G Q G Y G Met T E A** G P V L T Met S L A F A
K Q P F E V K P G G C G T V V R N A E L K I V D P E S G A S L P R N Q P **G E I C**
I R G D Q I Met K G Y L N D P E S T R T T I D K E G W L H T G D I G F I D D D D
E L F I V D R L K E L I K Y K G F Q V A P A E L E A L L I T H P S V S D A A V V P
Met K D E A A G E V P V A F V V R S N N S Q L T E D E V K Q F I S K Q V V F Y K
R I N R V F F I E A I P K S P S G K I L R K D L R A K L A A G F P N Stop

Figure 58. Full-length amino acid sequence of Sa4CL1. The highly conserved peptide motifs are underlined.

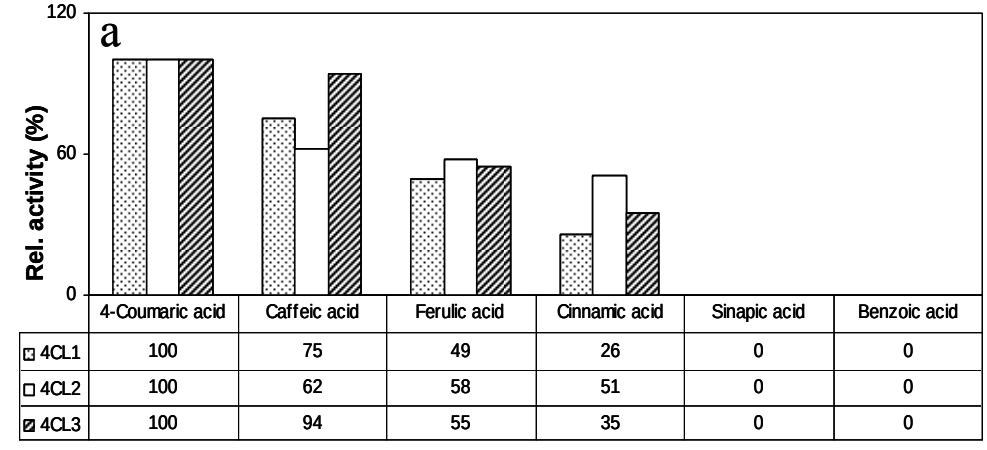
A second highly conserved peptide motif found in all members of the superfamily of adenylate-forming enzymes is Box I. This motif is rich in Gly, Ser and Thr and contains near the C-terminus an absolutely conserved Lys [Figure 58]. These structural features form a so-called phosphate-binding loop. In this flexible loop, substitution of the corresponding Lys^{211} in the SSGTTGLPK^{211}G motif of AT4CL2 with Ser has no effect on the K_m values for both ATP and caffeic acid. The mutant Lys211Ser exhibited only 3% of wild-type activity. Accordingly, it appears questionable whether the SSGTTGLPKG signature motif of AT4CL2 is directly involved in nucleotide binding. A similar conclusion was drawn from the crystal structure of the phenylalanine-activating subunit of gramicidin S synthetase 1 (PheA) in a complex with AMP and phenylalanine (Conti et al., 1997). In this conformation, this conserved signature motif forms a flexible loop between two β-strands and was suggested to accommodate the pyrophosphate leaving group during catalysis rather than direct participation in ATP binding. The recombinant enzymes of *4CL* genes in plants often demonstrate distinct substrate utilization profiles within a species (Hu et al., 1998; Ehlting et al., 1999). Generally, 4-coumamrate and caffeate are the best substrates, followed by ferulate. Cinnamate is a very poor substrate (Douglas, 1996). So far, none of the 4CLs examined activated hydroxyl- or aminobenzoic acids, for example in *Ruta graveolens* (Endler et al., 2008), *Centaurium erythraea* (Barillas and Beerhues, 1997), and *R. ideaus* (Kumar and Ellis, 2003), to mention a few. This result was also confirmed for Sa4CLs. A distinct benzoate-CoA-ligase was partially purified from *Clarkia breweri* (Beuerle and Pichersky, 2002), which also activated anthranilate with 50% efficiency but did not accept (hydroxy)cinnamic acids. 4CL activity toward sinapate has been described in few classes of angiosperms (Knobloch and Hahlbrock, 1975; Kutsuki et al., 1981). Only one heterologously expressed 4CL isoform from *Glycin max* (Gm4CL1) and two from *A. thaliana* (At4CL4 and At4CL5) have so far been described with such activity (Lindermayr et al., 2002; Hamberger and Hahlbrock, 2004; Costa et al., 2005). Whether sinapyl alcohol is generated from sinapate *via* sinapoyl:CoA has been of interest because 4CL activity toward sinapate is quite low in many angiosperms (Lee et al., 1995, 1997; Lee and Douglas, 1996; Allina et al., 1998; Hu et al., 1998; Ehlting et al., 1999; Hamada et al., 2003). Instead, studies have suggested alternative pathways for sinapyl alcohol biosynthesis, in which 5-hydroxylation and subsequent *O*-methylation of the guaiacyl nucleus can occur through cinnamyl aldehyde or the alcohol pathway in poplar, aspen, etc. (Matsui et al., 1994; Meyermans et al., 2000). Aspen 4CL (Pt4CL1) selectively converts caffeate to its CoA ester in its reaction with the mixture of cinnamate derivatives (Harding et al., 2002). This is also persuasive evidence that syringyl monolignol is synthesized through coniferyl

aldehyde or coniferyl alcohol, but not through sinapate in plants. However, feeding experiments with shoots of black locust (Hamada et al., 2004) showed that deuterium-labeled sinapate was incorporated into syringyl lignin. This incorporation was also observed in oleander (*Nerium indicum*) but not in magnolia (*Magnolia kobus*) or *A. thaliana*. These results suggest that sinapate is a precursor of sinapyl alcohol in black locust and oleander and that syringyl lignin biosynthesis in angiosperms probably varies, depending on the species (Yamauchi et al., 2002, 2003). In soybean, a single amino acid deletion determines whether or not 4CL can use sinapic acid as a substrate (Gm4CL2 and Gm4CL3; Lindermayr et al., 2003). Based on structural data obtained by homology modeling of gramicidin S synthetase with *A. thaliana At*4CL2 as a test case, Schneider et al. (2003) had combined structural, biochemical, and mutational analyses. The study revealed that the substrate-binding pocket (SBP) of this and all other 4CLs with the same specificity-determining amino acid code is too small to accommodate sinapate. The structural model additionally indicated that correct orientation of sinapic acid in the At4CL2 SBP is prevented by the amino acid residues Val^{355} and Leu^{356}, which interfere with the 5-methoxy group of the substrate. Directed mutations enlarging this pocket to a size similar to that realized in *Gm*4CL1 (an enzyme capable of activating sinapic acid) indeed led to the expected broadening of the substrate specificity, including the conversion of sinapate. *In silico* studies in the same context identified one additional, functionally undefined ”4CL-like” protein in *A. thaliana* that contained a similar large SBP and was a sinapate-activating 4CL. Accessibility of the At4CL2 SBP for monomethoxylated and dimethoxylated cinnamic acid derivatives is regulated mainly by size exclusion, whereas cinnamic acid conversion is controlled by the overall hydrophobicity of the At4CL2 SBP (Schneider et al., 2003). As previously reported, ferulic acid activation by At4CL2 depends on the presence of a small amino acid residue at either position 293 or 320, whereas bulky residues at both positions prevent ferulic acid activation by steric interference with its 3-methoxy group (Stuible and Kombrink, 2001). Cinnamate:CoA ligase was postulated to particularly carry a hydrophobic residue at the position corresponding to Asn-256 in At4CL2 (Schneider et al. 2003), which, however, is not the case with the cinnamate-preferring Ri4CL2 isoenzyme (Kumar and Ellis, 2003).

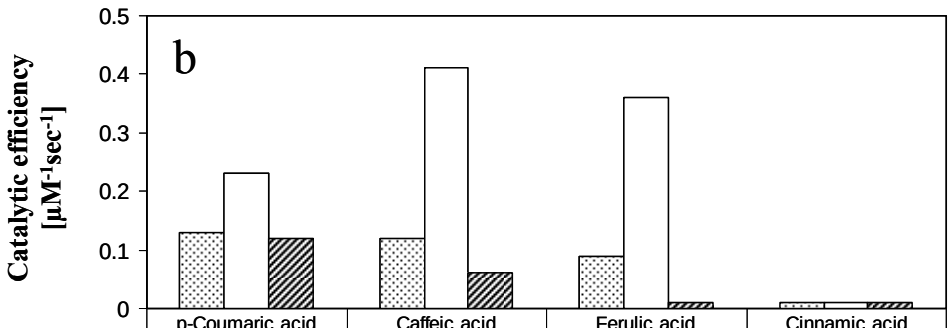
A. thaliana At4CL1 and At4CL2 show remarkable differences in their substrate utilization profiles. Identification of two adjacent domains involved in substrate recognition was accomplished by employing a domain swapping approach to generate several chimeric proteins that contain parts of the At4CL1 and At4CL2 amino acid sequences. The enzymatic properties of the recombinant chimeric enzymes were then determined using ferulate and *p*-

coumarate as substrates. This analysis revealed the existence of two small contiguous domains within the 4CL protein that by themselves have profound effects on 4CL substrate utilization, and thus clearly define 4CL substrate recognition domains (Ehlting et al., 2001).

In *Arabidopsis*, cinnamic acid is a poor substrate for all three 4CLs isoforms and sinapic acid (3,5-methoxy-4-hydroxycinnamic acid) is not activated at all (Ehlting et al., 1999). The same is true for Sa4CLs [Figure 59]. For *Sorbus* spp, absence of reports recording accumulation of B-ring unsubstituted flavonoids or benzoate derivatives, whose biosynthesis might be predicted to require cinnamoyl-CoA as a precursor, supports the poor utilization of cinnamic acid by Sa4CLs. It has been reported before that some members of the 4CL family may be susceptible to artifactual acceptance of cinnamate as a substrate when expressed as recombinant forms (Lee and Douglas, 1996). In *Sorbus*, the highest catalytic efficiencies for caffeate and ferulate were observed with Sa4CL2 and Sa4CL1 (Table 4), which confirms their role in lignin biosynthesis. None of the recombinant proteins had detectable activities against sinapate, or against benzoic acid. Similar results had been reported with recombinant 4CLs from *R. idaeus* (Kumar and Ellis, 2003).



	4-Coumaric acid	Caffeic acid	Ferulic acid	Cinnamic acid	Sinapic acid	Benzoic acid
4CL1	100	75	49	26	0	0
4CL2	100	62	58	51	0	0
4CL3	100	94	55	35	0	0



	p-Coumaric acid	Caffeic acid	Ferulic acid	Cinnamic acid
4CL1	0.13	0.12	0.09	0.01
4CL2	0.23	0.41	0.36	0.01
4CL3	0.12	0.06	0.01	0.01

Figure 59. Substrate utilization profiles of Sa4CL isoforms represented as relative activity (a), and catalytic efficiencies (b).

Table 4. Steady-state kinetic parameters of the three Sa4CLs

Isozyme	Substrate	K_m [μM]	V_{max} [nkat/mg protein]	K_{cat} [sec^{-1}]	K_{cat}/K_m [$\mu M^{-1}sec^{-1}$]
Sa4CL1 Mw = 59.5*	*p*-coumaric	7.3	16.0	0.96	0.13
	caffeic	6.8	13.0	0.78	0.12
	ferulic	5.4	8.0	0.49	0.09
	cinnamic	21.0	4.5	0.26	0.01
Sa4CL2 Mw = 60.1* (Scharnhop, 2008)	*p*-coumaric	2.5	9.6	0.58	0.23
	caffeic	0.9	6.0	0.36	0.41
	ferulic	0.9	5.6	0.34	0.36
	cinnamic	23.0	4.9	0.30	0.01
Sa4CL3 Mw = 66.2* (Ramadan, 2006)	*p*-coumaric	9.0	16.0	1.00	0.12
	caffeic	12.5	12.0	0.80	0.06
	ferulic	50.0	5.7	0.38	0.01
	cinnamic	33.0	5.6	0.37	0.01

*Predicted molecular mass from amino acid sequence.

The coding region of Sa4CL1 showed 71 and 63% nucleotide sequence identity to Sa4CL2 and Sa4CL3, respectively. The derived amino acid sequences shared greatest homology to the corresponding 4CLs from *R. idaeus*, with decreasing homology to plant and bacterial *4CL-like* genes, luciferase, long-chain fatty acid-CoA ligase, and peptide synthetases.

Sa4CL3 (Ramadan, 2006), which belongs to Class II, contains a N-terminal extension of 49 amino acids, which is absent from Sa4CL1 and Sa4CL2. Extended N-termini were also found in *Lithospermum erythrorhizon* 4CL2, rice 4CL2, aspen 4CL2, *R. idaeus* 4CL3, and *Ruta graveolens* 4CL1, all of which are class II members (Yazaki et al., 1995; Hu et al., 1998; Kumar and Ellis, 2003; Endler et al., 2008). It has been proposed that these variant N-terminal regions might be involved in phenolic substrate binding specificity (Hu et al., 1998) but this has yet to be confirmed experimentally.

The transcript levels of the three Sa4CL isoenzymes detected in *S. aucuparia* cell cultures were low in the absence of elicitor but they all increased after the addition of chitosan. The changes in the mRNA amounts resembled the time course of accumulation of BIS1 mRNA which reached a maximum after 6 h, as observed previously on RNA blots (Liu et al., 2007). This coordinate regulation might suggest involvement of Sa4CLs in formation of cinnamoyl-CoA and/or benzoyl-CoA. However, the expression rates of all three *Sa4CL* genes were significantly lower than the transcription levels of the *BIS1* and *PAL* genes, questioning participation of Sa4CLs in benzoyl-CoA biosynthesis. Transcriptional activation of *4CL* genes was previously found after elicitor treatment of cell cultures from various plant species

including parsley, soybean and *Arabidopsis* (Douglas et al., 1987; Trezzini et al., 1993; Lindermayr et al., 2002). Expression of *4CL* genes was also induced in differentiated plants of parsley, potato, soybean and *Arabidopsis* following inoculation with different types of pathogens or exposure to other kinds of environmental stress such as wounding (Schmelzer et al., 1989; Becker-André et al., 1991; Uhlmann and Ebel, 1993; Lee et al., 1995). Under these conditions, phenylpropanoid derivatives play an important defensive role (Dixon and Paiva, 1995). They also protect from damage by UV-irradiation which strongly induces Sa4CL3 transcripts but poorly activates expression of the *Sa4CL1* and *Sa4CL2* genes. This differential regulation was earlier reported for the *4CL* genes present in other species, such as *Arabidopsis* (Ehlting et al., 1999).

Transgenic plants with down-regulated 4CL expression (e.g. tobacco, *Arabidopsis*, and poplar) show a reduction in lignin content with only moderate or no changes in composition (Kajita et al., 1996; Lee et al., 1997; Hu et al., 1999). 4CL is thus an unlikely candidate for manipulating lignin composition, at least in dicotyledons. This may be due to broad substrate utilization by the 4CLs involved in monolignol biosynthesis or by occurrence of methoxylation and hydroxylation of monolignol precursors at a later stage in the biosynthetic pathway. However, 4CL remains a likely candidate for modifying lignin content in plants.

The variation in expression patterns of the three Sa4CLs (Scharnhop, 2008) combined with the substrate utilization profiles of the three recombinant proteins strongly suggests that the different *Sa4CLs* gene family members have different functional roles during plant growth and development. Defining these roles will require creation and characterization of transgenic *Sorbus spp.* lines in which each gene has specifically been silenced.

Besides low and no activity with cinnamic and sinapic acids, respectively, Sa4CL3 converts ferulic acid less efficiently (0.01 $\mu M^{-1}sec^{-1}$) than do Sa4CL1 and Sa4CL2. Based on the signature motif of substrate specificity-determining amino acids (Stuible et al., 2000; Ehlting et al., 2001; Schneider et al., 2003), this isoform may offer a good candidate for mutational analysis to generate a novel substrate preference.

V. Summary

- Benzoic acid biosynthesis in plants is poorly understood. In cell cultures of *Sorbus aucuparia* (Rosaceae), the activities of the benzoic acid biosynthetic enzymes were co-ordinately stimulated by chitosan treatment, resulting in accumulation of the biphenyl phytoalexin aucuparin. The carbon skeleton of this inducible defence compound is formed by biphenyl synthase (BIS) from benzoyl-CoA and three molecules of malonyl-CoA.

- The formation of benzoyl-CoA proceeds *via* benzaldehyde as an intermediate. Benzaldehyde dehydrogenase (BD) catalyzes the last reaction of benzoic acid biosynthesis and converts benzaldehyde to benzoic acid. The preferred substrate for BD was benzaldehyde (K_m = 49 µM). Cinnamaldehyde and hydroxybenzaldehydes were relatively poor substrates. BD activity was dependent on the presence of NAD^+ as a cofactor (K_m = 67 µM). The pH and temperature optima were 9.5 and 40°C, respectively. NAD^+ was the preferred cofactor. The enzyme was inhibited by divalent cations, with Cu^{2+} and Zn^{2+} being most inhibitory.

- Benzoic acid is activated by CoA ligase. Using a homology-based approach, a cDNA encoding CoA ligase was cloned from elicitor-treated *S. aucuparia* cell cultures. The expressed protein was not soluble as His_6-tagged protein but functionally expressed as GST-fusion protein in *E. coli*. The 1644 bp ORF encoded a 59.5 kDa protein of 547 amino acids. The enzyme shared 84% identity with *Rubus idaeus p*-coumarate-CoA ligase 1 (4CL1) and 71 and 63.5% identities with Sa4CL2 and Sa4CL3, respectively, that were previously cloned from *S. aucuparia* cell cultures (Ramadan, 2006; Scharnhop, 2008). The new CoA ligase activated *p*-coumaric, caffeic, ferulic and cinnamic acids and was inactive with sinapic acid. Using radioisotopic assays, the lack of affinity for benzoic acid was demonstrated. The enzyme was thus identified as 4CL isoenzyme (Sa4CL1). Its temperature and pH optima were 37°C and 7.5, respectively. The highest catalytic efficiencies (K_{cat}/K_m) were observed with *p*-coumaric and caffeic acids. Sa4CL1 activity was strictly dependent on the presence of a divalent cation, preferably Mg^{2+}.

- After elicitor treatment of *S. aucuparia* cell cultures, the transcript levels of all three Sa4CLs increased, however, their maximum mRNA levels after 6-9h were significantly lower than the expression rates observed for the L-phenylalanine ammonia-lyase (*PAL*) and biphenyl synthase 1 (*BIS1*) genes. Sa4CL3 and PAL transcripts also accumulated in response to light treatment, whereas *BIS1* expression was not affected by irradiation.

- A fragment of a fourth cDNA encoding CoA ligase was isolated from chitosan-treated cell cultures. The PCR product of about 780 bp consisted of the 5´-promoter and untranslated region (483 bp) and the coding region including the start codon (297 bp). It shared 81% nucleotide sequence identity with the corresponding fragment of the Sa4CL1 cDNA. The functional behavior of this new CoA ligase remains to be resolved following full-length cloning.

- A full-length cDNA encoding aldehyde dehydrogenase (ALDH) was isolated from chitosan-treated cell cultures. The 1512 bp ORF encoded a 54.8 kDa protein of 503 amino acids. The enzyme shared 79-52% identity with ALDHs from castor, maize, tobacco, and snapdragon. RT-PCR analysis revealed that SaALDH transcripts were up-regulated upon elicitation, strongly suggesting its involvement in benzoic acid metabolism.

- A full-length cDNA encoding PAL with an ORF of 2160 bp was also cloned from elicitor-treated cell cultures. The 720 amino acid sequence with a predicted molecular mass of 78 kDa shared highest homology (97-88%) with PALs from *Pyrus* sp., *Prunus avium* and *R. idaeus*. Up-regulation of PAL transcripts in *S. aucuparia* cell cultures in response to elicitation was demonstrated.

VI. References

Abd El-Mawla AMA, Beerhues L; Benzoic acid biosynthesis in cell cultures of *Hypericum androsaemum***.** *Planta* **214:** 727-733 **(2002)**

Abd El-Mawla AMA, Schmidt W, Beerhues L; Cinnamic acid is a precursor of benzoic acids in cell cultures of *Hypericum androsaemum* L. but not in cell cultures of *Centaurium erythraea* RAFN. *Planta* **212:** 288-293 **(2001)**

Achnine L, Blancaflor EB, Rasmussen S, Dixon RA; Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell* **16:** 3098-3109 **(2004)**

Allina SM, Pri-Hadash A, Theilmann DA, Ellis BE, Douglas CJ; 4- Coumarate:coenzyme A ligase in hybrid poplar. Properties of native enzymes, cDNA cloning, and analysis of recombinant enzymes. *Plant Physiol* **116:** 743-754 **(1998)**

Altenschmidt U, Oswald B, Fuchs G; Purification and characterization of benzoate-coenzyme A ligase and 2-aminobenzoate-coenzyme A ligases from a denitrifying. *Pseudomonas sp*. *J Bacteriol* **173:** 5494-5501 **(1991)**

Andreoli C, Prokisch H, Hortnage IK, Mueller JC, Munsterkotter M, Scharfe C, Meitinger T; MitoP2, an integrated database on mitochondrial proteins in yeast and man. *Nuc Acid Res* **32:** D459-D462 **(2004)**

Austin MB, Noel JP; The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep* **20:** 79-110 **(2003)**

Barillas W, Beerhues L; 3-Hydroxybenzoate: coenzyme A ligase and 4-coumarate: coenzyme A ligase from cultured cells of *Centaurium erythraea*.*Planta* **202:** 112-116 **(1997)**

Barillas W, Beerhues L; 3-Hydroxybenzoate:Coenzyme A Ligase from Cell Cultures of *Centaurium erythraea*: Isolation and Characterization. *Biol Chem* **381:** 155-160 **(2000)**

Becker-Andre´ M, Schulze-Lefert P, Hahlbrock K; Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. *J Biol Chem* **266:** 8551-8559 **(1991)**

Beerhues L, Liu B, Raeth T, Klundt T, Beuerle T, Bocola M; Benzoic acid-specific type III polyketide synthases. *American Chemical Society Symposium Series 955* (**Rimando, A.M., Baerson, S.R., eds.)** 97-108 **(2007)**

Beerhues L, Liu B; Biosynthesis of biphenyls and benzophenones – Evolution of benzoic acid – specific type III polyketide synthases in plants. *Phytochemistry* **70:** 1719-1727 **(2009)**

Berger F ; Handbuch der Drogenkunde, Bd III, Maudrich, Wien **(1952)**

Bernards MA , Lewis NG; Alkyl ferulates in wound healing potato tubers. *Phytochemistry* **31:** 3409-3412 **(1992)**

Beuerle T, Pichersky E; Purification and characterization of benzoate:coenzyme A ligase from *Clarkia breweri*. *Arch Biochem Biophys* **400:** 258-264 **(2002)**

Bimboim HC, Doly J; A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7:** 1513-1523 **(1979)**

Borejsza-Wysocki W, Lester C, Attygalle AB, Hrazdina G; Elicited cell suspension cultures of apple (Malus x domestica) cv. Liberty produce biphenyl phytoalexins**.** *Phytochemistry* **50:** 231–235 **(1999)**

Bradford MM; A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72:** 248-254 **(1976)**

Claros MG, Vincens P ; Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* **241:** 779-786 **(1996)**

Chakraborty D, Sircar D, Mitra A. Phenylalanine ammonia-lyase-mediated biosynthesis of 2-hydroxy-4-methoxybenzaldehyde in roots of *Hemidesmus indicus*. *J Plant Physiol* **165:** 1033-40 **(2008)**

Chevallier A; The Encyclopedia of Medicinal Plants. Dorling Kindersley London ISBN 9-780751-303148 **(1996)**

Chiej R; Encyclopaedia of Medicinal Plants. MacDonald ISBN 0-356-10541-5 **(1984)**

Cohen SN, Chang ACY, Hsu I, Nonchromosomal antibiotic resistance in bacteria: Genetic transformation of *Escherichia coli* by R-factor DNA. *Proc. Natl. Acad. Sci*. USA **69:** 2110-2114 **(1972)**

Conti E, Franks NP, Brick P; Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure* **4:** 287-298 **(1996)**

Conti E, Stachelhaus T, Marahiel MA, Brick P; Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *Embo J* **16:** 4174-4183 **(1997)**

Costa MA, Bedgar DL, Moinuddin SGA, Kim K-W, Cardenas CL, Cochrane FC, Shockey JM, Helms GM, Amakura Y, Takahashi H, Milhollan JK, Davin LB, Browse J, Lewis NG; Characterization in vitro and in vivo of the putative multigene 4-coumarate:CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation. *Phytochemistry* **66:** 2072-2091**(2005)**

Crawford DL, Sutherland JB, Pometto ALIII, Miller JM; Production of an aromatic aldehyde oxidase by *Streptomyces viridosporus. Arch Microbiol* **131:** 351-355 **(1982)**

Croteau R, Kutchan TM, Lewis NG; Natural products (secondary metabolites). In: Buchanan BB, Gruissem W, Jones RL (ed), Biochemistry and Molecular Biology of Plants. USA, *American Society of Plant Physiologists* 1250-1318 **(2000)**

Cui X, Wise RP, Schnable PS; The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize. *Science* **272:** 1334-1336 **(1996)**

Cukovic D, Ehlting J, VanZiffle JA, Douglas CJ; Structure and evolution of 4-coumarate: coenzyme A ligase (4CL) gene families. *Biol Chem* **382:** 645-654 **(2001)**

Dagert M, Erlich D; Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene* **6 :**23-28 **(1979)**

de Groot H, Rauen U; Tissue injury by reactive oxygen species and the protective effects of flavonoids. *Fundam Clin Pharmacol* **12:** 249-55 **(1998)**

De Luca V, St Pierre B; The cell and developmental biology of alkaloid biosynthesis. *Trends Plant Sci* **5:** 1360-1385 **(2000)**

Dewick PM; Medicinal natural products. A biosynthetic approach, In: John Willy and Sons, Chichester **(1997)**

Dixon RA, Pavia NL; Stress-induced phenylpropanoid metabolism. *Plant Cell* **7:** 1085-1097 **(1995)**

Douglas CJ; Phenylpropanoid metabolism and lignin biosynthesis: from weeds to trees. *Trends plant Sci* **1:** 171-178 **(1996)**

Douglas C, Hoffmann H, Schulz W, Hahlbrock K; Structure and elicitor or u.v.-light-stimulated expression of two 4-coumarate:CoA ligase genes in parsley. *EMBO J* **6:** 1189-1195 **(1987)**

Egland PG, Gibson J, Harwood CS; Benzoate-coenzyme A ligase, encoded by badA, is one of three ligases able to catalyze benzoyl-coenzyme A formation during anaerobic growth of Rhodopseudomonas palustris on benzoate. *J Bacteriol* **177:** 6545-6551 **(1995)**

Ehlting J, Buttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E; Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J* **19:** 9-20 **(1999)**

Ehlting J, Shin JJ, Douglas CJ; Identification of 4-coumarate:coenzyme A ligase (4CL) substrate recognition domains. *Plant J* **27:** 455-465 **(2001)**

Emanuelsson O, Brunak S, von Heijne G, Nielsen H; Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2:** 953-971 **(2007)**

Endler A, Martens S, Wellmann F, Matern U; Unusually divergent 4-coumarate:CoA-ligases from *Ruta graveolens* L. *Plant Mol Biol* **67:** 335-346 **(2008)**

Enyedi AJ, Yalpani N, Silverman P, Raskin I; Localization, conjugation, and function of salicylic acid in tobacco during the hypersensitive reaction to tobacco mosaic virus. *Proc Natl Acad Sci* USA **89:** 2480-2484 **(1992)**

Erdtman H, Eriksson G, Norin T; Aucuparin and methoxyaucuparin, two phenolic biphenyl derivatives from the heartwood of *Sorbus aucuparia L. Acta Chem Scand* **17:** 1151-1156 **(1963)**

Farres J, Wang X, Takahashi K, Cunningham SJ, Wang TT, Weiner H; Effects of changing glutamate 487 to lysine in rat and human liver mitochondrial aldehyde dehydrogenase. *J Biol Chem* **269:** 13854-13860 **(1994)**

Farres J,Wang TTY, Cunningham SJ,Weiner H; Investigation of the active site cysteine residue of rat liver mitochondrial aldehyde dehydrogenase by site-directed mutagenesis. *Biochemistry* **34:** 2592-2598 **(1995)**

Ferrer J-L, Jez JM, Bowman ME, Dixon RA, Noel JP; Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis**.** *Nature Structural Biology* **6:**775-784 **(1999)**

Fischbach MA, Walsh CT; Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106:** 3468-3496 **(2006)**

Fischer TC, Gosch C, Pfeiffer J, Halbwirth H, Halle C, Stich K, Forkmann G; Flavonoid genes of pear (*Pyrus communis*). *Trees* **21:**521-529 **(2007)**

French CJ, Vance CP, Towers GHN; Conversion of *p*-coumaric acid to p-hydroxybenzoic acid by cell free extracts of potato tubers and *Polyporus hispidus*. *Phytochemistry* **15:** 564-566. **(1976)**

Frohne D, Pfänder HJ; poisonous plants. London ISBN 1-874545-94-4**.** English translation of the second (1982) edition of "Giftpflanzen" **(1984)**

Fulda M, Heinz E, Wolter FP; The fadD gene of *Escherichia coli* K12 is located close to *rnd* at 39.6 min of the chromosomal-map and is a new member of the AMP binding protein family. *Mol Gen Genet* **242:** 241-249 **(1994)**

Gaid MM, Sircar D, Beuerle T, Mitra A, Beerhues L; Benzaldehyde dehydrogenase from chitosan-treated *Sorbus aucuparia* cell cultures. *J Plant Physiol* **166:** 1343-1349 **(2009)**

Görlach J, Scmid J; Introducing Stu I sites improves vectors for the expression of fusion proteins with factor Xa cleavage sites. *Gene* **170:** 145-146 **(1996)**

Grieve ; A Modern Herbal. Penguin ISBN 0-14-046-440-9 **(1984)**

Gross GG, Zenk MH; Darstellung und Eigenschaften von Coenzyme A-Thioestern substituierter Zimtsa¨uren. *Z. Naturforsch* **21b:** 688-690 **(1966)**

Gulick AM, Starai VJ, Horswill AR, Homick KM, Escalante-Semerena JC; The 1.75 A crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry* **42:** 2866-2873 **(2003)**

Guo Y, Ma L, Ji Y, Pu G, Liu B, Du Z, Li G, Ye H, Wang H; Isolation of the 5'-end of plant gene by TATA-box Degenerate Primers polymerase chain reaction. *Mol Biotech* DOI 10.1007/s12033-010-9323-0 in press **(2010)**

Hahlbrock K, Scheel D; Physiology and molecular biology of phenylpropanoid metabolism. *Annu Rev Plant Physiol Plant Mol Biol* **40:** 347-469 **(1989)**

Häkkinen SH, Kärenlampi SO, Heinonen M, Mykkänen HM, Törrönen AR; Content of the Flavonols Quercetin, Myricetin, and Kaempferol in 25 Edible Berries. *J Agric Food Chem* **47:** 2274-2279 **(1999)**

Hamada K, Nishida T, Yamauchi K, Fukushima K, Kondo R, Tsutsumi Y; 4-Coumarate: coenzyme A ligase in black locust (*Robinia pseudoacacia*) catalyses the conversion of sinapate to sinapoyl-CoA**.** *J Plant Res* **117:** 303-310 **(2004)**

Hamada K, Ysutsumi Y, Nishida T; Treatment of poplar callus with ferulic and sinapic acids. II: Lignin biosynthesis and its related enzyme. *J Wood Sci* **49:** 366-370 **(2003b)**

Hamada K, Ysutsumi Y, Yamauchi K, Fukushima K, Nishida T; Treatment of poplar callus with ferulic and sinapic acids I: Incorporation and enhancement of lignin biosynthesis. *J Wood Sci* **49:** 333-338 **(2003a)**

Hamberger B, Hahlbrock K; The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proc Natl Acad Sci U S A* **101:** 2209-2214 **(2004)**

Harborne JB; The comparative biochemistry of phytoalexin induction in plants. *Biochemistry Systematics and Ecology* **27:** 335-367 **(1999)**

Harborne JB; The Flavonoids: Advances in Research since 1980. *New York: Chapman and Hall* **ISBN-**10: 0412287706 **(1988)**

Harding SA, Leshkeich J, Chiang VL, Tsai CJ; Differential substrate inhibition couples kinetically distinct 4-coumarate: coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. *Plant Physiol* **128:** 428-438 **(2002)**

Heath R, McInnes R, Lidgett A, Huxley H, Lynch D, Jones E,Mahoney N, Spangenberg G; Isolation and characterisation of three 4-coumarate : CoA-ligase homologue cDNAs from perennial ryegrass (*Lolium perenne*). *J Plant Physiol* **159:** 773-779 **(2002)**

Hirano JI, Miyamoto K, Ohta H; Purification and characterization of aldehyde dehydrogenase with a broad substrate specificity originated from 2 phenylethanol-assimilating *Brevibacterium sp*. KU1309. *Appl Microbiol Biotechnol* **76:** 357-363 **(2007)**

Holton TA , Cornish EC; Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* **7:** 1071-1083 **(1995)**

Hopwood DA, Sherman DH; Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet* **24:** 37-66 **(1990)**

Hrazdina G, Borejsza-Wysocki W, Lester C; Phytoalexin production in an apple cultivar resistant to *Venturia inaequalis*. *Phytopathology* **87:** 868-876 (**1997)**

Hrazdina G; Response of scab-susceptible (McIntosh) and scab-resistant (Liberty) apple tissues to treatment with yeast extract and *Venturia inaequalis*. *Phytochemistry* **64:** 485-492 **(2003)**

Hu LH, Sim KY; Cytotoxic polyprenylated benzoylphloroglucinol derivatives with an unusual adamantyl skeleton from *Hypericum sampsonii* (Guttiferae). *Org Lett* **1:**879-882 **(1999)**

Hu WJ, Kawaoka A, Tsai CJ, Lung J, Osakabe K, Ebinuma H, Chiang VL; Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). *Proc Natl Acad SciUSA* **95:** 5407-5412 **(1998)**

Huang W, Ma X, Wang Q, Gao Y, Xue Y, Niu X, Yu G, Liu Y; Significant improvement of stress tolerance in tobacco plants by overexpressing a stress-responsive aldehyde dehydrogenase gene from maize (*Zea mays*). *Plant Mol Biol* **68:** 451-463 **(2008)**

Hukkanen AT, Pölönen S S, Kärenlampi S O, Kokko H I; Antioxidant Capacity and Phenolic Content of Sweet Rowanberries. *J Agric Food Chem* **54 :** 112-119 **(2006)**

Hüttner C, Beuerle T, Scharnhop H, Ernst L, Beerhues L; Differential effect of elicitors on biphenyl and dibenzofuran formation in *Sorbus aucuparia* cell cultures. *J Agric Food Chem* accepted **(2010)**

Ibdah M, Chen Y-T, Wilkerson CG, Pichersky E; An Aldehyde Oxidase in developing Seeds of Arabidopsis converts benzaldehyde to benzoic acid. *Plant Physiol* **150(1):** 416-423 **(2009)**

Jarvis AP, Schaaf O, Oldham NJ; 3-Hydroxy-3-phenylpropanoic acid is an intermediate in the biosynthesis of benzoic acid and salicylic acid but benzaldehyde is not. *Planta* **212:** 119-126 **(2000)**

Kähkönen MP, Hopia AI , Vuorela HJ , Rauha JP, Pihlaja K, Kujala TS, Heinonen M ; Antioxidant Activity of Plant Extracts Containing Phenolic Compounds. *J Agric Food Chem* **47:** 3954-3962 **(1999)**

Kähkönen MP, Hopia AI, Heinonen M; Berry Phenolics and Their Antioxidant Activity. *J Agric Food Chem* **49**: 4076-4082 **(2001)**

Kajita S, Katayama Y, Omori S; Alterations in the biosynthesis of lignin in transgenic plants with chimeric genes for 4-coumarate: coenzyme A ligase. *Plant Cell Physiol* **37:** 957-965 **(1996)**

Kauss H, Franke R, Krause K, Conrath U, Jeblick W, Grimmig B, Matern U; Conditioning of parsley (*Petroselinum crispum L.*) suspension cells increases elicitor-induced incorporation of cell wall phenolics. *Plant Physiol* **102:** 459-466 **(1993)**

Kirch HH, Bartels D, Wei Y, Schnable PS, Wood AJ; The ALDH gene superfamily of *Arabidopsis*. *Trends Plant Sci* **9:** 371-377**(2004)**

Kirch HH, Schlingensiepen S, Kotchoni S, Sunkar R, Bartels D; Detailed expression analysis of selected genes of the aldehyde dehydrogenase (ALDH) gene superfamily in *Arabidopsis thaliana*. *Plant Mol Biol* **57:** 315-332 **(2005)**

Kitson TM;The effect of disulfiram on the aldehyde dehydrogenase of sheep liver. *Biochem J* **151:** 407-412 **(1975)**

Klessig DF, Malamy J; The Salicylic acid signal in plants. *Plant Mol Biol* **26:** 1439-1458 **(1994)**

Kliebenstein D, D'Auria J, Behere A, Kim J, Gunderson K, Breen J, Lee G, Gershenzon J, Last R, Jander G; Characterization of seed-specific benzoyloxyglucosinolate mutations in *Arabidopsis thaliana*. *Plant J* **51:** 1062-1076 **(2007)**

Klundt T, Bocola M, Lütge M, Beuerle T, Liu B, Beerhues L; A single amino acid substitution converts benzophenone synthase into phenylpyrone synthase. *J Biol Chem* **284:** 30957-30964 **(2009)**

Knobloch KH, Hahlbrock K; 4-coumarate:CoA ligase from cell suspension cultures of *Petroselinum hortense* Hoffm. *Arch Biochem Biophys* **184:** 237-248 **(1977)**

Knobloch KH, Hahlbrock K; Isoenzymes of p-coumarate: CoA ligase from cell suspension cultures of *Glycine max*. *Eur J Biochem* **52:** 311-320 **(1975)**

Koiwai H, Akaba S, Seo M, Komano T, Koshiba T; Functional expression of two Arabidopsis aldehyde oxidases in the yeast *Pichia pastoris*. *J Biochem* **127:** 659-664 **(2000)**

Kokubun T, Harborne JB; Phytoalexin induction in the sapwood of plants of the Maloideae (Rosaceae): biphenyls or dibenzofurans. *Phytochemistry* **40:** 1649-1654 **(1995)**

Kokubun T, Harborne JB, Eagles J, Waterman PG; Antifungal biphenyl compounds are the phytoalexins of the sapwood of *Sorbus aucuparia*. *Phytochemistry* **40:** 57-59 **(1995)**

Kokubun T, Harborne JB; A survey of phytoalexin induction in leaves of the Rosaceae by copper ions. *Z Naturforsch* **49C:** 628-634 **(1994)**

Komaraiah P, Naga Amrutha R, Kavi Kishor PB, Ramakrishna SV; Elicitor enhanced production of plumbagin in suspension cultures of *Plumbago rosea* L. *Enz Microbial Technol* **31:** 634-639 **(2002)**

Koshiba T, Saito E, Ono N, Yamamoto N, Sato M; Purification and properties of flavin- and molybdenum containing aldehyde oxidase from coleoptiles of maize. *Plant Physiol* **110:** 781-789 **(1996)**

Kotchoni SO, Bartels D; Water stress induces the up-regulation of a specific set of genes in plants: aldehyde dehydrogenase as an example. *Bulg J Plant Physiol* **Special Issue:** 37-51 **(2003)**

Kotchoni SO, Kuhns C, Ditzer A, Kirch HH, Bartels D; Over-expression of different aldehyde dehydrogenase genes in *Arabidopsis thaliana* confers tolerance to abiotic stress and protects plants against lipid peroxidation and oxidative stress. *Plant Cell Environ* **29:** 1033-1048 **(2006)**

Kumar A, Ellis BE; 4-coumarate:CoA ligase gene family in *Rubus idaeus*: cDNA structures, evolution, and expression. *Plant Mol Biol* **51:** 327-340 **(2003)**

Kutsuki H, Shimada M, Higuchi T; Distribution and roles of p-hydroxycinnamate: CoA ligase in lignin biosynthesis. *Phytochemistry* **21:** 267-271 **(1981)**

Lacombe E, Hawkins S, VanDoorsselaere J, Piquemal J, Goffner D, Poeydomenge O, Boudet, AM, GrimaPettenati J; Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: Cloning, expression and phylogenetic relationships. *Plant J* **11:** 429-441 **(1997)**

Laemmli UK; Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680-685 **(1970)**

Large AT, Connock MJ; Hydrogen peroxide generating alcohol and aldehyde oxidases in the giant African land snail, *Achatina fidica. J Exp Zool* **270:** 445-450 **(1994)**

Lee D, Douglas CJ; Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins. *Plant Physiol* **112:** 193-205 **(1996)**

Lee D, Ellard M, Wanner LA, Davis KR, Douglas CJ; The *Arabidopsis thaliana* 4-coumarate:CoA ligase (4CL) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA. *Plant Mol Biol* **28:** 871-884 **(1995)**

Lee D, Meyer K, Chapple C, Douglas CJ; Antisense suppression of 4-coumarate:coenzyme A ligase activity in *Arabidopsis* leads to altered lignin subunit composition. *Plant Cell* **9:** 1985-1998 **(1997)**

Lewis NG ; A 20th century roller coaster ride: a short account of lignification. *Curr Opin Plant Biol* **2:** 153-162 **(1999)**

Lewis NG, Yamamoto E; Lignin: Occurence, biogenesis and biodegradation. *Annu Rev Plant Physiol Plant MoI Biol* **41:** 455-496 **(1990)**

Lindahl R, Petersen DR; Lipid aldehyde oxidation as a physiological role for class 3 aldehyde dehydrogenases. *Biochem Pharmacol* **41:** 1583-1587 **(1991)**

Lindahl R; Aldehyde dehydrogenases and their role in carcinogenesis. *Crit Rev Biochem Mol Biol* **27:** 283-335 **(1992)**

Lindermayr C, Fliegmann J, Ebel J; Deletion of a single amino acid residue from different 4-coumarate:CoA ligases from soybean results in the generation of new substrate specificities. *J Biol Chem* **278:** 2781-2786 **(2003)**

Lindermayr C, Möllers B, Fliegmann J, Uhlmann A, Lottspeich F, Meimberg H, Ebel J; Divergent members of a soybean (*Glycine max* L.) 4-coumarate:coenzyme A ligase gene family. *Eur J Biochem* **269:** 1304-1315 **(2002)**

Linsmaier EM, Skoog F; Organic growth factor requirements of tobacco tissue cultures. *Physiol Plantarum* **18:** 100-127 **(1965)**

Liu B, Beuerle T, Klundt T, Beerhues L; Biphenyl synthase from yeast-extracttreated cell cultures of *Sorbus aucuparia*. *Planta* **218:** 492-496 **(2004)**

Liu B, Raeth T, Beuerle T, Beerhues L; A novel 4-hydroxycoumarin biosynthetic pathway. *Plant Mol Biol* **72:**17-25 **(2010)**

Liu B, Raeth T, Beuerle T, Beerhues L; Biphenyl synthase, a novel type III polyketide synthase. *Planta* **225:** 1495-1503 **(2007)**

Liu F, Cui X, Horner HT, Weiner H, Schnable PS; Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. *Plant Cell* **13:** 1063-1078 **(2001)**

Liu F, Schnable PS; Functional specialization of maize mitochondrial aldehyde dehydrogenases. *Plant Physiol* **130:** 1657-1674 **(2002)**

Liu ZJ, Sun YJ, Rose J, Chung YJ, Hsiao CD, Chang W-R, Kuo I, Perozich J,Lindahl R, Hempel J, Wang B-C; The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nature Structural Biology* **4:** 317-326 **(1997)**

Long MC, Nagegowda DA, Yasuhisa K, Ho KK, Kish CM, Schnepp J, Sherman D, Weiner H, Rhodes D, Dudareva N; Involvement of snapdragon benzaldehyde dehydrogenase in benzoic acid biosynthesis. *Plant J* **59:** 256-265 **(2009)**

Lòpez Barragàn MJ, Carmona M, Zamarro MT, Thiele B, Boll M, Fuchs G, Garcia JL, Diaz E; The bzd gene cluster, coding for anaerobic benzoate catabolism, in *Azoarcus* sp. strain CIB. *J Bacteriol* **186:** 5762-5774 **(2004)**

Löscher R, Heide L; Biosynthesis of *p*-hydroxybenzoate from *p*-coumarate and *p*-coumaroyl-coenzyme A in cell free extracts of *Lithospermum erythrorhizon* cell cultures**.** *Plant Physiol* **106:** 271-279 **(1994)**

Lozoya E, Hoffmann H, Douglas CJ, Schulz W, Scheel D, Hahlbrock K; Primary structures and catalytic properties of isoenzymes encoded by two 4-coumarate:CoA ligase genes in parsley. *Eur J Biochem* **176:** 661-667 **(1988)**

Makkar HPS, Siddhuraju P, Becker K; Methods in molecular biology, Plant secondary metabolites. Humana press New Jersey ISBN-10: 1-58829-993-7 **(2007)**

Malinowski J, Krzymowska M, Godon K, Hennig J, Podstolski A; A new catalytic activity from tobacco converting 2-coumaric acid to salicylic aldehyde. *Physiol Plantarum* **129:** 461-471 **(2007)**

Mandel M, Higa A; Calcium-dependent bacteriophage DNA infection. *J Mol Biol* **53:** 159-162 **(1970)**

Marchitti SA, Brocker C, Stagos D, Vasiliou V; Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin Drug Metab Toxicol* **4:** 697-720 **(2008)**

Matsui N, Fukushima K, Kamada K, Nishikawa Y, Yasuda S (1994); On the behavior of monolignol glucosides. I. Synthesis of monolignol glucosides labeled with ^{2}H at the hydroxymethyl group of side chain, and polymerization of the labeled monolignol in vitro. *Holzforschung* **48:** 215-221**(1994)**

May JJ, Kessler N, Marahiel MA, Stubbs MT; Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A* **99:** 12120-12125 **(2002)**

Me´traux JP, Signer H, Ryals J, Ward E, Wyss-Benz M, Gaudin J, Raschdorf K, Schmid E, Blum W, Inverardi B; Increase in salicylic acid at the onset of systemic acquired resistance in cucumber. *Science* **250:** 1004-1006 **(1990)**

Meyermans H, Morreel K, Lapierre C, Pollet B, De Bruyn A, Busson R, Herdewijn P, Devreese B, Beeumen JV, Marita JM, Ralph J, Chen C, Burggraeve B, Montagu MV, Messens E, Boerjan*a* W; Modifications in Lignin and Accumulation of Phenolic Glucosides in Poplar Xylem upon Down-regulation of Caffeoyl-Coenzyme A *O*-Methyltransferase, an Enzyme Involved in Lignin Biosynthesis. *J Biol Chem* **275:** 36899-36909 **(2000)**

Mishra PK, Drueckhammer DG; Coenzyme A analogues and derivatives: Synthesis and applications as mechanistic probes of coenzyme A ester-utilizing enzymes. *Chem Rev* **100:** 3284-3309 **(2000)**

Mitra A, Mayer MJ, Mellon FA, Michael AJ, Narbad A, Parr AJ, Waldron K,Walton NJ; 4-Hydroxycinnamoyl-CoA-hydratase/lyase, an enzyme of phenylpropanoid cleavage from *Pseudomonas*, causes formation of C6-C1 glucose conjugates when expressed in hairy roots of *Datura stramonium. Planta* **15:** 79-89 **(2002)**

Miyakodo M, Watanabe K, Ohno N, Nonaka F, Morita A; Isolation and structural determination of eriobofuran, a new dibenzofuran phytoalexin from leaves of loquat, *Eriobotrya japonica L. J Pesticide Sci* **10:** 101-106 **(1985)**

Moore BS, Hertweck C; Biosynthesis of marine natural products: microorganisms (Part A). *Nat Prod Rep* **22:** 580-593 **(2005)**

Moreno PRH, Van der Heijden R, Verpoorte R; Elicitor-mediated induction of isochorismate synthase and accumulation of 2, 3-dihydroxy benzoic acid in *Catharanthus roseus* cell suspension and shoot cultures. *Plant Cell Rep* **14:** 188-191**(1994)**

Mustafa NR, Verpoorte R; Chorismate derived C6C1 compounds in plants. *Planta* **222**: 1-5 **(2005)**

Nair RB, Bastress KL, Ruegger MO, Denault JW, Chapple C; The Arabidopsis thaliana REDUCED EPIDERMAL FLUORESCENCE1 gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell* **16:** 544-554 **(2004)**

Negre F, Kish CM, Boatright J, Underwood B, Shibuya K, Wagner C, Clark DG, Dudareva N; Regulation of methylbenzoate emission after pollination in snapdragon and petunia flowers. *Plant Cell* **15:** 2992-3006 **(2003)**

Omarov R, Akaba S, Koshiba T, Lips H; Aldehyde oxidase in roots, leaves and seeds of barley (*Hordeum vulgare* L.). *J Exp Bot* **50:** 63-69 **(1999)**

op den Camp RGL, Kuhlemeier C; Aldehyde dehydrogenase in tobacco pollen. *Plant Mol Biol* **35:** 355-365 **(1997)**

Ossowski S, Schwab R, Weigel D; Gene silencing in plants using artificial micro-RNAs and other small RNAs. *Plant J* **53:** 674-690 **(2008)**

Perozich J, Wang B, Lindahl R, Hempel J; Relationships within the aldehyde dehydrogenase extended family. *Protein Sci* **8:** 137-146 **(1999)**

Pitta-Alvarez SI, Giulietti AM; Influence of chitosan, acetic acid and citric acid on growth and tropane alkaloid production in transformed roots of *Brugmansia candida*. Effect of medium, pH and growth phase. *Plant Cell Tiss Org Cult* **59:**31-38 **(1999)**

Podstolski A, Havin-Frenkel D, Malinowski J, Blount JW, Kourteva G, Dixon RA; Unusual 4-hydroxybenzaldehyde synthase activity from tissue cultures of the vanilla orchid *Vanilla planifolia*. *Phytochemistry* **61**: 611-620 **(2002)**

Potter D, Eriksson T, Evans RC, Oh S, Smedmark JEE, Morgan DE, Kerr M, Robertson KR, Arsenault M, Dickinson TA, Campbell CS; Phylogeny and classification of Rosaceae. *Pl Syst Evol* **266:** 5-43 **(2007)**

Ramadan H; Molecular analysis of coenzyme A ligase from benzoate-metabolizing *Sorbus aucuparia* cell cultures. Dissertation TU Braunschweig, Braunschweig **(2006)**

Rasmussen S, Dixon RA; Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into phenylpropanoid pathway. *Plant Cell* **11:** 1537-1552 **(1999)**

Ribnicky DM, Shulaev V, Raskin I; Intermediates of salicylic acid biosynthesis in tobacco. *Plant Physiol* **118**: 565-572 **(1998)**

Rice-Evans CA, Miller NJ, Paganga G; Structure-antioxidant activity relationships of flavonoids and phenolic acids. *Free Radic Biol Med* **20:** 933-956 **(1996)**

Rodriguez-Zavala JS, Weiner H; Structural aspects of aldehyde dehydrogenase that influence dimmer-tertamer formation. *Biochemistry* **41:** 8229-8237 (**2002)**

Ross AF; Systemic acquired resistance induced by localized virus infections in plants. *Virology* **14:** 340-358 **(1961)**

Rout UK, Weiner H; Involvement of serine 74 in the enzyme-coenzyme interaction of rat liver mitochondrial aldehyde dehydrogenase. *Biochemistry* **33:** 8955-8961 **(1994)**

Ryals JA, Neuenschwander UH, Willits MG, Molina A, Steiner H-Y, Hunt MD; Systemic acquired resistance**.** *Plant Cell* **8:** 1809-1819 **(1996)**

Sambrook J, MacCallum P; Molecular cloning: a laboratory manual (3rd ed.). New York:Cold Spring Harbor Laboratory Press **(2001)**

Sanger F, Nicklen S, Coulson AR; DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74:** 5463-5467 **(1977)**

Sarkanen KV, Ludwig CH ; Lignin, Occurrence, Formation, Structure and Reactions. *Wiley/Interscience, New York* 95-240 **(1971)**

Scharnhop H; Untersuchungen zur Biosynthese aromatischer Sekundärmetabolite in Zellkulturen von *Sorbus aucuparia* L. und *Centaurium erythraea* RAFN. Dissertation TU Braunschweig, Braunschweig **(2008)**

Schauenstein E, Esterbauer H, Zollner H; Aldehydes in Biological Systems: Their Natural Occurrence and Biological Activities. *Pion*, London **(1977)**

Schmelzer E, Krüger-Lebus S, Hahlbrock K; Temporal and spatial patterns of gene expression around sites of attempted fungal infection in parsley leaves. *Plant Cell* **1:** 993-1001 **(1989)**

Schneider K, Hövel K, Witzel K, Hamberger B, Schomburg D, Kombrink E, Stuible HP; The substrate specificity-determining amino acid code of 4-coumarate: CoA ligase. *PNAS* **100:** 8601-8606 **(2003)**

Schnitzler JP, Madlung J, Rose A, Seitz HU; Biosynthesis of *p*-hydroxybenzoic acid in elicitor-treated carrot cell cultures. *Planta* **188:** 594-600 **(1992)**

Schramm G, Bruchhaus I, Roeder T; A simple and reliable 5'-RACE approach. *Nucleic Acids Res* **28:** 96e **(2000)**

Schröder G, Brown JWS , Schröder J; Molecular analysis of resveratrol synthase: cDNA, genomic clones and relationship with chalcone synthase. *Eur J Biochem*. **172:** 161-169 **(1988)**

Schröder J, Raiber S, Berger T, Schmidt A, Schmidt J, Soares-Sello AM, Bardshiri E, Strack D, Simpson TJ, Veit M, Schröder G; Plant polyketide synthases: a chalcone synthase-type enzyme which performs a condensation reaction with methylmalonyl-CoA in the biosynthesis of C-methylated chalcones. *Biochem* **37(23):** 8417–8425 **(1998**)

Schröder J; Comprehensive Natural Products Chemistry. **Sankawa U**. *Elsevier Science: Amsterdam* **1:** 749-771 **(1999)**

Schühle K, Gescher J, Feil U, Paul M, Jahn M, Schagger H,Fuchs G; Benzoate-coenzyme A ligase from *Thauera aromatica*: an enzyme acting in anaerobic and aerobic pathways. *J Bacteriol* **185:** 4920-4929 (**2003)**

Scotto-Lavino E, Du G, Frohman MA; Amplification of 5' end cDNA with 'new RACE'. *Nat Protoc* **1:** 3056-3061 **(2006)**

Seo M, Koiwai H, Ahaba S, Komano T, Oritani T, Kamiya T, Koshiba T; Abscisic aldehyde oxidase in leaves of *Arabidopsis thaliana*. *Plant J* **23:** 481-488 **(2000)**

Serino L, Reimmann C, Baur H, Beyeler M, Visca P, Haas D; Structural genes for salicylate biosynthesis from chorismate in *Pseudomonas aeruginosa*. *Mol Gen Genet* **249:** 217-228 **(1995)**

Shahmuradov IA, Gammerman AJ, Hancock JM, Bramley PM, Solovyev VV; PlantProm: a database of plant promoter sequences. *Nucleic Acids Res* **31:** 114-117 **(2003)**

Shulaev V, Silverman P, Raskin I; Airborne signalling by methyl salicylate in plant pathogen resistance. *Nature* **385:** 718-721 **(1997)**

Sircar D, Mitra A; Accumulation of *p*-hydroxybenzoic acid in hairy roots of *Daucus carota* 2: Confirming biosynthetic steps through feeding of inhibitors and precursors. *J Plant Physiol* **166:** 1370-1380 **(2009)**

Sircar D, Mitra A; Evidence for *p*-hydroxybenzoate formation involving enzymatic phenylpropanoid side chain cleavage in hairy root of *Daucus carota*. *J Plant Physiol* **165:** 407- 414 **(2008)**

Sircar D; Enzymatic route to 4-hydroxybenzoic acid formation in hairy root cultures of *Daucus carota* L. Dissertation Indian Institute of technology Kharagpur **(2009)**

Smith DA; Toxicity of phytoalexins. In Phytoalexins, **Bailey JA, Mansfield JW.** *New York: John Wiley and Sons* 218-252 **(1982)**

Smith DB, Johnson KS; Single-step purification of polyketides expressed in *Escherichia coli* as fusion with glutathione S-transferase. *Gene* **67:** 31-40 **(1988)**

Sophos NA, Vasiliou V; Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chem Biol Interact* **143-144:** 5-22 **(2003)**

Stachelhaus T, Mootz HD, Marahiel MA; The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6:** 493-505 **(1999)**

Staunton J, Weissman KJ; Polyketide biosynthesis: a millennium review**.** *Nat Prod Rep* **18:** 380-416 **(2001)**

Stein T, Vater J, Kruft V, Otto A, Wittmann-Liebold B, Franke P, Panico M, Mcdowell R, and Morris HR; The Multiple Carrier Model of Nonribosomal Peptid Biosynthesis at Modular Multienzymatic Templates. *J Biol Chem* **271:** 15428-15435 **(1996)**

Stöckigt J, Zenk, MH; Chemical syntheses and properties of hydroxycinnamoyl-coenzyme A derivatives. *Z Naturforsch* **30c:** 352–358 **(1975)**

St-Pierre B, De Luca V; In Recent Advances in Phytochemistry Evolution of Metabolic Pathways, Romeo JT, Ibrahim R, Varin L, De Luca V. Elsevier, Oxford. Vol. **34:** 285-317 **(2000)**

Stuible HP, Buttner D, Ehlting J, Hahlbrock K, Kombrink E; Mutational analysis of 4-coumarate:CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. *FEBS Lett* **467:** 117-122 **(2000)**

Stuible HP, Kombrink E; Identification of the Substrate Specificity-conferring Amino Acid Residues of 4-Coumarate:Coenzyme A Ligase Allows the Rational Design of Mutant Enzymes with New Catalytic Properties. *J Biol Chem* **276:** 26893–26897 **(2001)**

Trezzini GF, Horrichs A, Somssich IE; Isolation of putative defense-related genes from *Arabidopsis thaliana* and expression in fungal elicitor-treated cells. Plant Mol Biol **21:** 385-389 **(1993)**

Tsuji H, Tsutsumi N, Sasaki T, Hirai A, Nakazono M; Organspecific expressions and chromosomal locations of two mitochondrial aldehyde dehydrogenase genes from rice (*Oryza sativa* L.), *ALDH2*a and *ALDH2b*. *Gene* **305:** 195-204 **(2003)**

Uhlmann A, Ebel J; Molecular cloning and expression of 4-coumarate:coenzyme A ligase, an enzyme involved in the resistance of soybean (*Glycine max*) against pathogen infection. *Plant Physiol* **102**: 1147-1156 **(1993)**

Van Moerkercke A, Schauvinhold I, Pichersky E, Haring MA, Schuurink RC; A plant thiolase involved in benzoic acid biosynthesis and volatile benzenoid production. *Plant J* **60:** 292-302 **(2009)**

Vasiliou V, Bairoch A, Tipton KF, Nebert DW; Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenetics* **9:** 421-434 **(1999)**

Villegas M, Brodelius PE; Elicitor-induced hydroxycinnamoyl-CoA:tyramine hydroxytransferase in plant cell suspension cultures**.** *Physiol plant* **78:** 414-420 **(1990)**

Wang C-Z, Maier UH, Eisenreich W, Adam P, Obersteiner I, Keil M, Bacher A, Zenk MH;Unexpected biosynthetic precursors of amarogentin-a retrobiosynthetic ^{13}C NMR study. *Eur J Org Chem* **8:** 1459-1465**(2001)**

Wang C-Z, Maier UH, Keil M, Zenk MH, Bacher A,Rohdich F,Eisenreich W; Phenylalanine-independent biosynthesis of 1,3,5,8-tetrahydroxycanthone.A retrobiosynthetic NMR study with root cultures of *Swertia chirata*. *Eur J Biochem* **270:** 2950-2958 **(2003)**

Wang X, Weiner H; Involvement of glutamate 268 in the active site of human liver mitochondrial (class 2) aldehyde dehydrogenase as probed by site-directed mutagenesis. *Biochemistry* **34:** 237–243 **(1995)**

Welle R, Grisebach H ; Phytoalexin synthesis in soybean cells: Elicitor induction of reductase involved in biosynthesis of 6'-deoxychalcone. *Arch Biochem Biophys* **272:** 97-102 **(1989)**

Werner I, Bacher A, Eisenreich W; Retro-biosynthetic NMR studies with ^{13}C-labeled glucose. Formation of gallic acid in plants and fungi. *J Biol Chem* **272:** 25474-25482 **(1997)**

Whetten R, Sederoff R; Lignin biosynthesis**.** *The Plant Cell* **7:** 1001-1013 **(1995)**

Widyastuti SM, Nonaka F, Watanabe K, Sako N, Tanaka K; Isolation and characterization of two aucuparin-related phytoalexins from *Photinia glabra. Maxim Ann Phytopath Soc Japan* **58:** 228-233 **(1992)**

Wildermuth MC, Dewdney J, Wu G, Ausubel FM; Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* **414:** 562-565 **(2001)**

Wildermuth MC; Variations on a theme: synthesis and modification of plant benzoic acids. *Curr Opin Plant Biol* **9:** 288-296 **(2006)**

Winkel-Shirley B; Biosynthesis of flavonoids and effects of stress. *Curr Opin Plant Biol* **5:** 218-223 **(2002)**

Wood AJ, Duff RJ. The aldehyde dehydrogenase (ALDH) gene superfamily of the moss Physcomitrella patens and the algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri. The Bryologist* **112:** 1-11 **(2009)**

Wu S, Chappell J; Metabolic engineering of natural products in plants; tools of the trade and challenges for the future. *Curr Opin Biotechnol* **19:** 145-152 **(2008)**

Yalpani N, Ledn J, Lawton MA, Raskin I; Pathway of salicylic acid biosynthesis in healthy and virus-inoculated tobacco. *Plant Physiol* **103:** 315-321 **(1993)**

Yamauchi K, Yasuda S, Fukushima K; Evidence for the biosynthetic pathway from sinapic acid to syringyl lignin using labeled sinapic acid with stable isotope at both methoxy groups in *Robinia pseudoacacia* and *Nerium indicum*. *J Agric Food Chem* **50:**3222-3227 **(2002)**

Yamauchi K, Yasuda S, Hamada K, Tsutsumi Y, Fukushima K; Multiform biosynthetic pathway of syringyl lignin in angiosperms. *Planta* **216:** 496-501 **(2003)**

Yasuhara A, Akiba-Goto M, Fujishiro K, Uchida H, Uwajima T, Aisaka K; Production of aldehyde oxidases by microorganisms and their enzymatic properties. *J Biosci Bioeng* **94:** 124-129 **(2002)**

Yazaki K, Heide L, Tabata M; Formation of *p*-hydroxybenzoic acid from p-coumaric acid by cell free extract of *Lithospermum erythrorhizon* cell cultures. *Phytochemistry* **30:** 2233-2236 (**1991)**

Yoshida A, Rzhetsky A, Hsu LC, Chang C; Human aldehyde dehydrogenase gene family. *Eur J Biochem* **251:** 549-557 **(1998)**

Zenk MH; Biosynthesis of vanillin in *Vanilla planifolia* Andr. *Z Pflanzenphysiol* **53:** 404-414 **(1965)**

Zheng C-F, Weiner H; Role of the highly conserved histidine residues in rat liver mitochondrial aldehyde dehydrogenase as studied by site-directed mutagenesis. *Arch Biochem Biophys* **305:** 460-466 **(1993)**

VII. Appendix

5´AAAGGAGGCTCCGAGGAAGGCGGAGAGTCGACGCACATTAGCTTAATGTCCTG
ACCCACCGCCACCATCCAAACCCACTTTCACCAAACCAAACCCCATGCACTCCTC
CCTTCCGCCTCCAGTCACCTATCACCACCAACCTCGAATGTGTCTTATCAAATTTC
CAATCTTTCACTATCACGGTTTCACCCCCAAACAATTCCATAATCCCATTTCCCTC
CCCCCCACCAACCACACCTCCAGTTGTATATGAAGAACCCCCATTGACCTCCCCT
CTTCCCTCAAGCACCAATTCAGCTCAAAATAAACCAACAATTTTGGTGTCTTTTTT
CTCATTCGATAATCCCATTTTTCCAAGAGATGGCTATAGAAACTATCCCAAACGA
CATCGTTTACCGGTCCAAACTCCCTGACATCCCCATCCCAAAACACCTCCCTCTC
CACTCCTACTGCCTCCACAACAAAAACCACTCGAGCTCTAAGCCCTGTATCATCG
ACGGCGCCACCGGAGACATATACACCTTCGCCGATGTGGAACTTAACGCGCGCA
GAGTCGCGTCAGGGCTCAACAAGCTTGGAATCCAACAAGGCGACGTCATCATGC
TCTTGCTCCCCAACTCCCCAGCGTTCGCCTTCGCCTTCCTCGGAGCCTCCTTTCGC
GGCGCTATGACAACCGCGGCGAACCCTTTCTTCACACCCGCCGAGATCTTAAAA
CAGGCCAAGGCCTCAAAAGCTAAACTCATCATCACCTTAGCGTGCTACTACGAC
AAGGTCAAGGACTTATCGTCATCAAGTGACGATGTTCACGACATTAAGCTAATGT
GCGTCGACTCTCCGCCTGATCCGAGCTGTTTGCATTTCTCCGAGCTTCTTCAAGCC
GATGAAAACGACATGCCGGAGGTTGACATCAGCCCAGACGACGTCGTCGCTCTA
CCCTACTCGTCCGGGACGACAGGCTTGCCTAAGGGGGTGATGCTAACGCACAAG
GGTCTTGTGACAAGCGTTGCTCAGCAGGTGGACGGGGAAAATCCCAACTTGTAT
TATAGCACCGACGACGTCGTCTTATGCGTGCTGCCACTTTTTCATATATATTCTTT
GAACTCGGTATTGCTTTGTGGACTTAGAGCCGGAGCTGCCATTTTGATGATGAAC
AAGTTTGAGATTGTTTCTCTGTTGGGGTTGATCGACAAGTACAAGGTTAGTATTG
CACCGATCGTGCCGCCGATAGTGTTGGCCATCGCCAAGTTTCCCGATCTTGATAA
GTACGATTTGTCGTCAATTCGAGTGCTTAAGTGTGGAGGGGCACCCCTTGGGAAG
GAGCTTGAGGATACTGTGAGAGCCAAGTTTCCCAATGTCACACTTGGTCAGGGA
TATGGAATGACAGAGGCAGGGCCAGTATTGACCATGTCATTGGCATTTGCCAAG
CAACCCTTCGAGGTCAAACCAGGTGGATGTGGCACCGTCGTCCGAAATGCAGAG
CTCAAGATCGTTGATCCTGAATCTGGTGCTTCTTTGCCACGCAACCAGCCTGGAG
AGATTTGCATTAGAGGTGACCAGATCATGAAAGGTTATCTTAATGATCCGGAGTC
GACAAGGACAACCATAGACAAGGAAGGTTGGCTACACACCGGTGATATAGGCTT
CATTGATGATGATGATGAGCTATTCATTGTTGATCGGTTGAAGGAACTGATCAAA
TACAAAGGATTTCAAGTGGCCCCTGCTGAACTTGAAGCCTTGCTCATCACCCATC
CTAGTGTTTCCGATGCTGCTGTTGTCCCAATGAAGGATGAGGCAGCTGGAGAGGT
TCCGGTTGCATTTGTAGTAAGGTCAAATAATTCTCAGCTCACTGAGGACGAAGTC
AAGCAATTTATCTCTAAACAGGTTGTATTTTACAAAAGAATAAATCGAGTATTTT
TCATCGAAGCCATTCCGAAGTCACCGTCGGGCAAAATCTTGCGGAAGGACTTGA
GAGCAAAGCTTGCTGCTGGGTTTCCAAACTGA 3´

Nucleotide sequence of the Sa4CL1 coding region (red) and the 5´UTR (black). The underlined sequences served as primers designed for expression experiments.

5′CTATAGGGCAAGCAGTGGTATCAACGCAGAGTGAGAGGACCAACGGAAACGG
CAGCCATGGGGAGTGATTTCAACGGCACCTCTGCTTCCGACTCCTTCGTCAAGAC
TCCCACAATAAAGTTCACCCAGCTGTTCATCAATGGCGAATTCCTCGATTCCGTT
TCAGGTAAAACATTTGAGACGATAGATCCAAGAACAGGGGACGTGGTAACCAGA
GTTGCAGAAGGAGACAAGGAAGACGTTGACTTGGCTGTCAAGGCGGCGCGTGCT
GCCTTCGACCATGGCTCTTGGCCTCGCTTGCCCGGCGCTGAGAGGGGAAGGATA
ATGATGAAGTTTGCAGACTTAATTGACAACATGTxAGAAGAACTAGCTATCTTGG
ATACTGTTGATGCCGGGAAGTTGTTCAGCATGGGCAAGACCATTGACATACCGC
AGGTAGCAGAAATGGTACGTTATTATGCAGGTGCAGCTGACAAAATCCATGGAG
AGGTGCTCAAAATGTCGCGCGAACTTCATGGTTATACGTTGCTTGAACCCATTGG
TGTTGTGGGGCTCATTGTTCCCTGGAATTTCCCGAGCACCCTGCTTTTCGGAAAG
GTCAGCCCTGCCTTAGCTGCTGGATGCACCATGGTCATCAAACCTGCTGAGCAAA
cACCTCTATCTGCAtTATACTATGCTCWTCTGGCTAAGTTGGCTGGTGTTCCTGAC
GGAGTGMTCAATGTCATAACTGGATTTGGAAAGACTGCTGGTGCTGCCATCAGC
TWTCATATGGACATTGACAAAGTTAGTTTTACTGGTTCCACCGAGGTAGGGCGCT
TAGTAATGCAGGCTGCAGCAAAGAGCAATCTGAAGCCAGTTTCACTTGAACTAG
GAGGCAAATCACCCCTTGTGATTTTTGACGACGCTGATATAAATATGGCTGCTGA
TCTTGCTCTCTTGGGAATCCTTTACAACAAGGGAGAAATTTGCGTGGCAAGTTCT
CGTGTTTATGTTCAAGAAGGGATTTATGATGAATTTGTGAAGAAGTTACAAGAG
AAGGCGAAGGATTGGGTAGTCGGGGATCCTTTTGATCCTAATGTCCGTCAAGGA
CCGCAGGTTGATAAGAAGCAGTTCGAAAAAATCCTATCCTACATTGAGCATGGG
AAGAAGGAAGGAGCCACATTGTTAACAGGGGGCAAGCCTGTGGGCAACAAGGG
ATATTACATTGAGCCTACAATATTTACTGATGTCAAGGATGACATGGTCATAGCC
CAGGATGAAATATTTGGACCAGTACTGGCGCTGATGAAGTTCAAGACAATCGAG
GAGGCGATACAGAGAGCCAACAACACCAGATACGGGCTAGCAGCAGGGATCAT
AACCAAGGACTTGAATGTGGCCAACACTGTCTCAAGGTCGATCCGCGCAGGCAT
TATTTGGATCAACTGCTACTTTGCTTTCGATCGAGACTGCCCTTACGGAGGGTAT
AAGATGAGCGGGTTCGGAAGAGACTTTGGTATGCAGGGCCTGTACCATTATCTC
CATACCAAATCTGTTGTCACTCCACTTTTTAACTCTCCCTGGCTCTAAATTGTGAT
ATAATCTTAAGGCTGGCAAATTGTGACCTTGTATTTGTGCAACATATACGTATAT
GTTATGTTGTATATCTGTTTCCTTTTGTTGTGGGAGATTTTATATTTTATTTACTGA
TGGAAAATACAGAAGTTTGAGTTTGAAATATAGTAGCAGAGGCTACAAATTATG
AAAAAAAAAAAA 3′

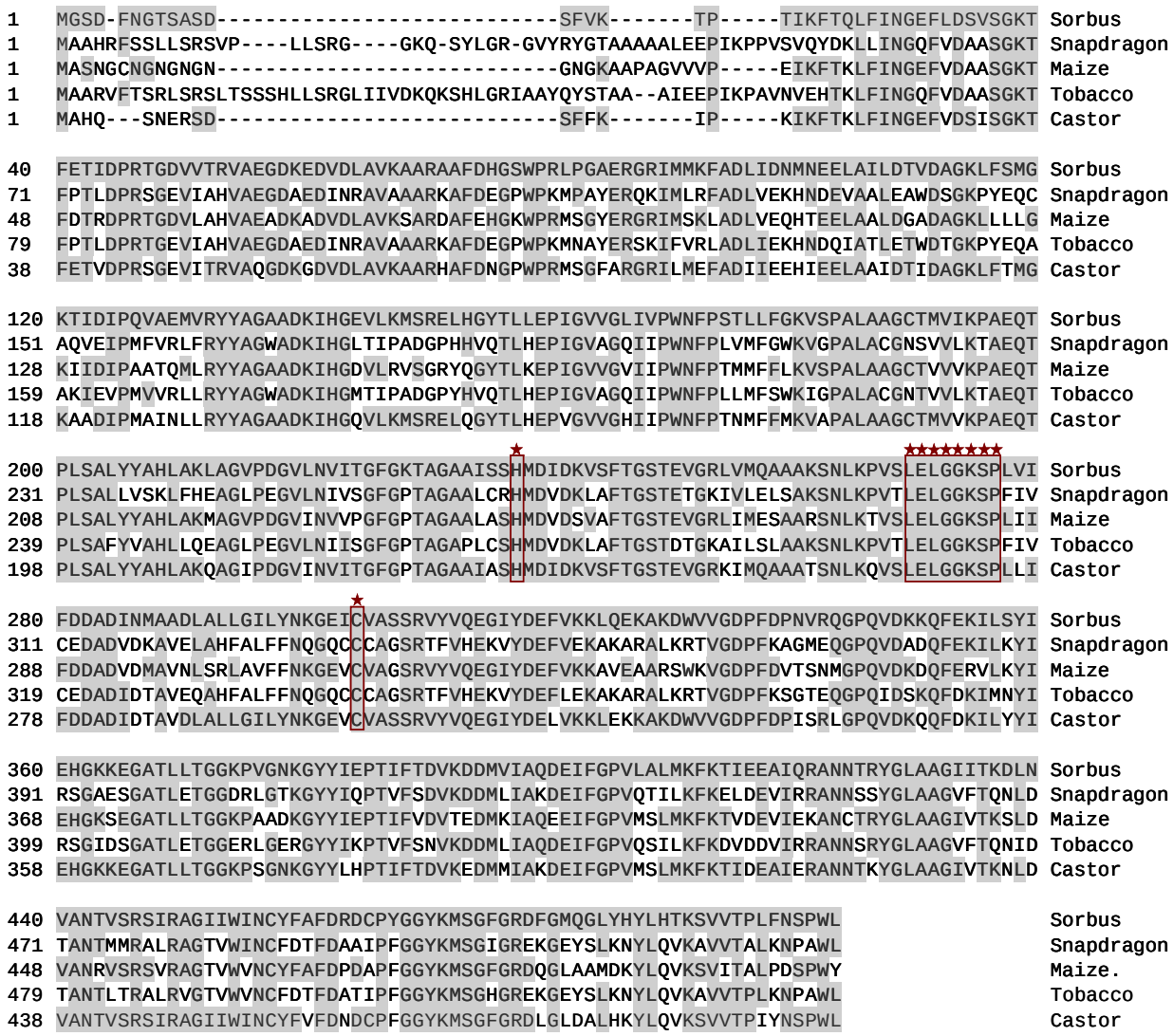
Nucleotide sequence of the ALDH coding region (red) and the 5′ and 3′ UTRs (black). The underlined sequences served as primers designed for expression experiments.

VII Appendix

5’ATGGAGGCGGAAACCATCACGCAAAATGGGAAAAACGGCCACCACCAGAA
CGGTGCTGTGGAGTCCCCGCTTTGCATTAAAAAGGACCCGTTGAACTGGGGT
CTAGCGGCGGATTCACTAAAAGGGAGCCACTTGGATGAAGTGAAGCGTATGG
TGGCGGAGTACAGAAAGCCGGTGGTGAAGCTCGGCGGAGAGAGCCTCACCA
TTTCCCAAGTTGCAGCCATAGCCACTCATGACACCGGGGTCAAGGTTGAGCT
CTCTGAGTCGGCCAGGGCTGGGGTCAAGGCCAGCAGTGATTGGGTCATGGAC
AGCATGGGCAAAGGGACTGACAGCTATGGTGTCACCACCGGGTTTGGTGCAA
CCTCCCACCGGAGAACAAAGCAAGGCGCTGCCCTTCAGAAGGAGCTAATTAG
ATTCTTGAACGCTGGAGTGTTTGGAAGTGCTACAGAATCAGGCCACACACTA
CCACATCAAGCAACTAGAGCAGCCATGTTGGTCAGAATCAAC<u>ACACTCCTCC</u>
<u>AAGGCTACTCC</u>GGCATAAGATTCGAAATCTTGGAAGCCATTACCAAGTTCTT
GAATAACAACGTCACTCCATGCTTGCCCCTACGGGGCACAATCACCGCCTCT
GGTGATCTTGTCCCGCTGTCCTACATTGCCGGTTTACTCACCGGCAGGCCAAA
CTCCAAAGCTGTCGGACCAAACGGTCAGACCCTCAATGCCTCTGAAGCTTTTG
AGCTAGTAGGGATCAATTCTGGGTTTTTCGAGTTGCAGCCTAAAGAATGGTTA
ACTCTTGTGAACGGCACTGCTGTTGGTTCTGGCTTGGCCTCCACGGTTCTTTTC
GAGACCAACATTTTGGCGTTGCTAGCGGAAATTTTGTCAACCTTCGCTGAAGT
GATGCAGGGGAAACCGGAGTTTACTGATCACTTGACGCACAAGTTGAAGCAC
CACCCTGGACAGATTGAGGCAGCTGCAATTATGGAACATATTTTGGATGGCA
GCTCTTATGTCAAAGCTGCTAAGAAGTTGCATGAGCAGGATCCCCTGCAGAA
GCCAAAGCAGGATCGCTACGCTCTCCGAACATCACCTCAGGGGTTAGGACCT
CAAATTGAAGTGATTCGTTATTCCACCAAATCCATTGGGAGGGGGATCAACT
CAGTTAACGACAACCCTTTGATTGATGTTTCGAGGAACAAGGCCTTGCACGG
AGGCAACTTCCAGGGGATCCCAATTGGAGTTTCCATGGACAACACTCGCTTG
GCGATTGCCTCCATTGGGAAGCTCATGTTTGCGCAATTTTCCGAGCTTGTCAA
TGACTTTTACAACAATGGTTTGCCTTCAAATCTGTCCGGAGGAAGGAATCCTA
GCTTGGATTACGGTTTCAAGGGGGCTGAAATCGCCATGGCATCTTATTGTTCC
GAGCTGCAGTTTCTCGCAAATCCGGTTACTAACCATGTCCAGAGTGCTGAGC
AGCACAACCAAGATGTTAACTCTTTGGGGTTGATCTCTTCAAGAAAGACAGC
TGAGGCGGTTGATATCTTGAAGCTCATGTCTTCTACATTTTT<u>GGTTGCGCTTTG</u>
<u>TCAGTCCAT</u>CGATTTGAGGCATTTGGAGGAGAACTTGAGGAACACTGTAAAG
AACATGGTGAGCCAAGTCGCCAAGAGGACTTTAACAACTGGGGTAAATGGG
GAGCTTCACCCCTCAAGATTCTGCGAGAAGGATCTGCTCAAAGTTGTCGATA
GGGAGTATGTTTTTGCCTACATTGATGACCCCTGCAGTGCCACTTATCCATTG
ATGCAGAAACTGAGGCAAGTGCTTGTTGAGCATGCTTTGACCAATGGTGAGA
GTGAGAAGAATGCAAGCACTTCGATCTTCCAAAAGATTGGAGCTTTCGAGGA
AGAGCTGAAAACGCTTTTGCCTAAAGAGGTGGAGAGTGCAAGGAGTGCAATT
GAGAGCGGAAATGCTGCAGTTCCAAACAGAATTGCGGAATGCAGGTCTTATC
CCTTGTACAAATTTGTGAGGGAGGAGTTGGGAGGAGAGTACCTGACCGGGGA
GAAGGTCAGGTCACCGGGCGAGGAGTGTGACAGGGTGTTCCAAGCTATCTGC
CAGCGAAAGATTATCGACCCGATTCTAGGTTGCCTCGAGGGTTGGAACGGCG
CACCTCTTCCTATCTGTTAG 3’

Nucleotide sequence of the SaPAL ORF. The underlined sequences served as primers designed for expression experiments

VII Appendix



Amino acid sequence alignment of various ALDHs. Conserved residues are indicated by asterisks.